

-2751-

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2670

- 5 A DNA sequence (GASx971) was identified in *S.pyogenes* <SEQ ID 7839> which encodes the amino acid sequence <SEQ ID 7840>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

10

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 2671

A DNA sequence (GASx972) was identified in *S.pyogenes* <SEQ ID 7841> which encodes the amino acid sequence <SEQ ID 7842>. Analysis of this protein sequence reveals the following:

Possible site: 46

25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3226(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35

Example 2672

A DNA sequence (GASx973) was identified in *S.pyogenes* <SEQ ID 7843> which encodes the amino acid sequence <SEQ ID 7844>. Analysis of this protein sequence reveals the following:

Possible site: 29

40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1830(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45

-2752-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2673

A DNA sequence (GASx975) was identified in *S.pyogenes* <SEQ ID 7845> which encodes the amino acid sequence <SEQ ID 7846>. Analysis of this protein sequence reveals the following:

Possible site: 45

10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.4757(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

20 >GP:BAB07248 GB:AP001519 unknown [Bacillus halodurans]
 Identities = 46/134 (34%), Positives = 73/134 (54%)

 Query: 23 KQPQDEKKYTDADVDAIIDKKFAKWKSEQEAEKSEAKKMAKMEKEKADYEKQKLLDELQ 82
 K + E+ +T +V+ I+ + A+ ++E EA+K+AKMN ++K +YE +KL E +
 Sbjct: 66 KPNKTERLFTQEEVNRIVKDRILARALKDKEEAIKEAEKLAKMNABQKREYELEKLRRENE 125

 Query: 83 ELKNDKTRNELTAVARQMFASEINVNDVGLVVTLDAEQTKANVTTLANAFKVIADD 142
 +LK + R EL A +M E+ I +DDVL VV DAEQT+ V T + K+
 Sbjct: 126 QLKKAQMRVELGREATKMLGEAGIMADDDVLSFVVRDDAEQTQEA VKTFISLVDKLADMR 185

 Query: 143 RKALVRQTTTPSTGG 156
 K ++ P G
 Sbjct: 186 MKEKLKGRPPPKDG 199

35 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2674

A DNA sequence (GASx976) was identified in *S.pyogenes* <SEQ ID 7847> which encodes the amino acid sequence <SEQ ID 7848>. Analysis of this protein sequence reveals the following:

Possible site: 24

40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2478(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

50 >GP:AAC79545 GB:U88974 ORF30 [Streptococcus thermophilus temperate
 bacteriophage 01205]

-2753-

Identities = 43/119 (36%), Positives = 66/119 (55%), Gaps = 16/119 (13%)

Query: 9 SKEILHNLDYEAISVTLDSENKIG-----KKVVPAGTILAGKDKSIFEDRKQKVETVTNEE 63
 + I+ +L Y+A+S T+DS+ G KK + AGT++AG SIF+DR + V
 5 Sbjct: 9 TSNIVRSLEPYKAVSATVDSSYPGVLVDGKKYIKAGTLVAGNGGSIFDDRTKSV----- 61

Query: 64 VSTKEYVDGILLTDVDTNGDAVGSCVYRGTTINADKLADSSVAENYDDLEEVLP HIVFI 122
 V K +GI+L DVDLT + V S +Y G + DK+ + D +++ LP + FI
 10 Sbjct: 62 VENKTEPEGIVLYDVDLTIDNTV-SVLYAGEVYKDKVNGGDIT--DITVKKALPLVKFI 116

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2675

A DNA sequence (GASx978) was identified in *S.pyogenes* <SEQ ID 7849> which encodes the amino acid sequence <SEQ ID 7850>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4238(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC79546 GB:U88974 ORF31 [Streptococcus thermophilus temperate bacteriophage O1205]
 Identities = 195/343 (56%), Positives = 256/343 (73%), Gaps = 1/343 (0%)

Query: 1 MALIHEIITSNIKGFYNKNNVENTLGEKAFPPKQQLGLKLSFIKGAAGKPVTLKAAA 60
 M LI++ +T+ NI G++NA ENV +TLGE FP ++QLG KLS+IKGA+G+ V LKAAA
 30 Sbjct: 1 MGLIYDKVTASNIAGYFNALQENVSSTLGESIFPARKQLGTKLSYIKGASGQSVALKAAA 60

Query: 61 FDTKVPLRDRMAVELIDEEMPPFFKEAMLVKEADRQQLNMLAQTKNNELIDTILASIYNDQ 120
 FDT V +RDR++ E+ DE+MPFFKEAMLVKE DRQQLN++ + N L++TI+A I+ND
 35 Sbjct: 61 FDTNVTIRDRVSAEMHDEQMPFFKEAMLVKENDRQQLNLVKDSGNVAVLNTIVAGIFNDN 120

Query: 121 ATLIAGAKARLEAMRMEVLKSGKIHQSNQVMDIDYGLAEDQTTKPDAKWDSAGTATPL 180
 TL+ GA+ARLEAMRM+VL+ GKI S+GV KDIDYG+ D + W G ATPL
 40 Sbjct: 121 LTLVNGARARLEAMRMQVLATGKIAFTSDGVNKKDIDYGVKPDHKKQVSKSWAEPG-ATPL 179

Query: 181 KDIEKAIKMAERGFVPEAIIMNSKTFSLIKNAESTLDVVKPMAENGAAVTKRDLNNTYLE 240
 D+E AIE E G PE +MN+KTF LI+ A ST+ V+KP+A +G+AVTK +L Y+
 45 Sbjct: 180 ADLEDAIETARELGLNPERAVMNAKTFGLIRKAASTVKVIKPLAGDGSVATKAELENYIA 239

Query: 241 DELQIKVILKDGFMFVGDDGESRKYPFDGATLVPNGNLGYTVFGTTPEQSDLLGGEATDA 300
 D + ++L++G + D GE K++PDG TL+PNG LG TVFGTTPE+SDL +A
 50 Sbjct: 240 DNFGVSIVLENGTYRNDKGEVSKFYPDGHLTLIPNGPLGNTVFGTTPEESDLFADNTVNA 299

Query: 301 NVSIVETGIAITTTTCTDPVNVQTKVSMIALPSFERLEEVIHII 343
 V IV+ GIA+TTTCTDPVNVQTKVSM+ALPSFERL++V+++
 Sbjct: 300 EVEIVDNGIAVTTTCTDPVNVQTKVSMVALPSFERLDDVYML 342

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2754-

Example 2676

A DNA sequence (GASx979) was identified in *S.pyogenes* <SEQ ID 7851> which encodes the amino acid sequence <SEQ ID 7852>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 46

   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3319(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

15 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2677

A DNA sequence (GASx980) was identified in *S.pyogenes* <SEQ ID 7853> which encodes the amino acid sequence <SEQ ID 7854>. Analysis of this protein sequence reveals the following:

```

20   Possible site: 55

   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
25      bacterial cytoplasm --- Certainty=0.2385(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

30 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAC34404 GB:AF158600 gp113 [Streptococcus thermophilus
    bacteriophage Sfi11]
    Identities = 53/109 (48%), Positives = 79/109 (71%), Gaps = 4/109 (3%)

35   Query: 11  IVKNVKLDLGIEDDNQDQLEMLLNRI TDHFKANYGVLEIDNAFSFVLEDCLIARFNRRG 70
      +++NV +DL I DDN  LL +LL RI +HFKA YGV E+D+  +F+ EDCL+ RFNRRG
    Sbjct: 9   VIQNVSVDLNINDDN---LLGILLERIVNHFKA EYGVDEVDDNLAFIFEDCLVKRFNRRG 65

    Query: 71  SERAKTEEVGHHKTTYDHLNEFEPYDAMIMAKLNLIKDKSRKGGLYFL 119
      +E A++E ++GH  +YYD+ NEF+PYD M+  +L      ++++G + FL
40   Sbjct: 66  AEGARSESIDGHSMYYDNENEFDPYDNMLQ-RLYGTSGQAKEGEVLFL 113

```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2678

A DNA sequence (GASx981) was identified in *S.pyogenes* <SEQ ID 7855> which encodes the amino acid sequence <SEQ ID 7856>. Analysis of this protein sequence reveals the following:

```

    Possible site: 49

50   >>> Seems to have no N-terminal signal sequence

```


-2755-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5714(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA59188 GB:X84706 b3 [Bacteriophage B1]

Identities = 28/82 (34%), Positives = 49/82 (59%), Gaps = 2/82 (2%)

10

Query: 1 MRVADRVTFVKTT-DEQYNPDLGEYTHTEVISITKPCFVMDMGMEKSVQIFGDYQKDRKV 59
 +RY D VTF+K + D Y+PDLGE+ E + D+G ++SV++FGD +K KV
 Sbjct: 1 LRYLDEVTFIKESPDSDHYDPLGEWVEKEPTRTVFSANITDIGTDRSVEVFGDIKKGAKV 60

15

Query: 60 IYLKQPYT-KAFDYCEYEGRRY 80
 + + + +DY E++ +++
 Sbjct: 61 MRMMPLFNMPKYDYIEFDNKKW 82

20

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2679

A DNA sequence (GASx982) was identified in *S.pyogenes* <SEQ ID 7857> which encodes the amino acid sequence <SEQ ID 7858>. Analysis of this protein sequence reveals the following:

Possible site: 14

25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2509(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC34406 GB:AF158600 gp114 [Streptococcus thermophilus
 bacteriophage Sfil11]

Identities = 44/103 (42%), Positives = 65/103 (62%), Gaps = 5/103 (4%)

35

Query: 17 GLKKKLELIKKDAVKK---IVRDNGTQLQRMINKAVFTKGYSTGATRRSITMQIGDGG 73
 GL + + ++K + +K ++R G++L+ +N+A F KGYSTGATRRSIT+Q+
 Sbjct: 8 GLDEMAQSLLKNASPEKRSKVLRYGSKLKEAAVNRAQFNKGYSTGATRRSITLQVESDK 67

40

Query: 74 LSVKVKPGTHYAGYLERGTRLMSKQPFVLPALKEQKVKFRKDL 116
 +V+ T Y+GYLE GTR M QPF+ PAL E K ++L

45

Sbjct: 68 ATVEAL--TSYSGYLEVGTRKMEAQPFMKPALDEVAPKMVEEL 108

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2680

A DNA sequence (GASx983) was identified in *S.pyogenes* <SEQ ID 7859> which encodes the amino acid sequence <SEQ ID 7860>. Analysis of this protein sequence reveals the following:

Possible site: 45

50

-2756-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3098(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:AAA32612 GB:L31366 putative [Bacteriophage Tuc2009]
 Identities = 88/129 (68%), Positives = 108/129 (83%)

 Query: 1 MIKTRDQSIFDEMFKRIQSLGFKVYDYKPMTEVPYPFVEMESTDAEYIPNKDDIKGSVEL 60
 MIKTRDQSIFDE+FKRIQ+LG+ VYDYKPM EV YPFVE+E+T + NK DIKG+V L
 15 Sbjct: 1 MIKTRDQSIFDELFKRIQALGYTTVYDYKPMNEVGYPFVELENTQTIHEANKTDIKGTVSL 60

 Query: 61 MLSVWGVQKKRKQVSDMASAIFSQALTVESSDVFRWSLNRQSSIQMLDDTTTTPPLKRA 120
 LSVWG+QKKRK+VSDMAS IF+QAL + ++D + W+LN++ S+IQMLDDTTT TPLKRA
 20 Sbjct: 61 SLSVWGLQKKRKEVSDMASNIFNQALNISATDGYSWALNSQASTIQMLDDTTTHTPLKRA 120

 Query: 121 IVTLRFNLR 129
 ++ L F LR
 Sbjct: 121 LINLEFRLR 129

25 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 antigens for vaccines or diagnostics.

Example 2681

A DNA sequence (GASx984R) was identified in *S.pyogenes* <SEQ ID 7861> which encodes the amino acid
 sequence <SEQ ID 7862>. Analysis of this protein sequence reveals the following:

30 Possible site: 36

 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.1736(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

40 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 antigens for vaccines or diagnostics.

Example 2682

45 A DNA sequence (GASx985) was identified in *S.pyogenes* <SEQ ID 7863> which encodes the amino acid
 sequence <SEQ ID 7864>. Analysis of this protein sequence reveals the following:

 Possible site: 27

 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 50 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2757-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

5  >GP:AAA32613 GB:L31366 structural protein [Bacteriophage Tuc2009]
    Identities = 81/185 (43%), Positives = 111/185 (59%), Gaps = 22/185 (11%)

    Query: 4  QLEAKQGIHSILLFRLIKEASSEATKLAFTQTEHEVGKSRDVGQKTKDGIIQSVGALEY 63
               +L AKQG ILL+RLI +A+ EAA KLAFTQTEH K+RD + TKDG I S+ A+EY
    10  Sbjct: 3  ELTAKQGGDIILLYRLLSKATKEAAWKLAFTQTEHSNEKTRDYNITATKDGITIGSLAAIEY 62

    Query: 64  DFKATSILAKGDVLAALKLEKAMENGELVEIWDIDLEETSNGDSDNKLANVWGIDKNGTN 123
               ATSI A GD +++KA ++GE+++W+ID E
    15  Sbjct: 63  SLSATSIANGDPHLDEMDKAFDDGEIIDVWEIDKAEG----- 101

    Query: 124  RGNKYLATYYQGYISSFSAKKNAEENIEIEMEFAINGVGQKGFATLTDAQKAAVQYAFK 183
               +GKY A Y + Y++SFS + N+E+ +E+ +EF + G QKG ATLT+ Q VQY FK
    20  Sbjct: 102 -SDGKYKAKYLRAYLTFSYEPNSEDALSLSEFGVFGKPKQKQATLTTEEQANVVQYVFK 160

    Query: 184  DTTKG 188
               DT G
    25  Sbjct: 161 DTVAG 165
  
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 2683

A DNA sequence (GASx986) was identified in *S.pyogenes* <SEQ ID 7865> which encodes the amino acid sequence <SEQ ID 7866>. Analysis of this protein sequence reveals the following:

```

    Possible site: 55

    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2273 (Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    35  Sbjct: 161 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
  
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

40  >GP:CAA59192 GB:X84706 a2 [Bacteriophage B1]
    Identities = 54/111 (48%), Positives = 72/111 (64%), Gaps = 1/111 (0%)

    Query: 1  MQLEIKGKTHNVKFGTRFVAEMDKNHIAERQGFQAGLQSSV-PFLIDHSVVTIAEVIY 59
               M+L IKGK + KFG +FV E+DKN + E+ G FG L + P L ++ TL+ V++
    45  Sbjct: 1  MELTIKKGQVHFQKFGVKFVRELDKNLVIEQNGVSFGLALAVKIIPELEMANIATLSNVLF 60

    Query: 60  TGTITEPPRPSLNDIYDYIDEVEDIEKLFDDVLDLQSNASKLFMAQVEK 110
               G TE P+ S DI D+IDE EDIEKLFDDVL E+ +SN KL A++ K
    50  Sbjct: 61  LGNRTETPKLSQGDIDDFIDECEDEIEKLFDDVLKEITESNTGKLIKAKMTK 111
  
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2684

A DNA sequence (GASx987) was identified in *S.pyogenes* <SEQ ID 7867> which encodes the amino acid sequence <SEQ ID 7868>. Analysis of this protein sequence reveals the following:

-2758-

Possible site: 36

>>> Seems to have no N-terminal signal sequence

5 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2735(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

10 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA59193 GB:X84706 c2 [Bacteriophage B1]
 Identities = 40/111 (36%), Positives = 57/111 (51%), Gaps = 10/111 (9%)

15 Query: 2 IVLNCIRYLGMTDINEIGRLTLYEYDLLMTGKALAAVDESHKAHQAWINHQVTATKLVG 61
 +++ +R G+ D++ R+T+ EY + L +DE ++QAW N QV ATK G
 Sbjct: 15 MMIRFLRCFGIQDLSVFERMTIREYSIRSIAFQLRTLDEEEFIYEQAWANWQVQATKQQG 74

20 Query: 62 GKKNKKEVPVYKFKDFFD---YEEEIRKI-TQEIDEGYDKKGM DLLKAN 108
 K P+Y FK FFD E EI I + E D K +DL+ KAN
 Sbjct: 75 KK-----PLYPTFKKFFDKKKLENEILGIESPENKFKDKNLIDLMMKKAN 119

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 **Example 2685**

A DNA sequence (GASx989) was identified in *S.pyogenes* <SEQ ID 7869> which encodes the amino acid sequence <SEQ ID 7870>. Analysis of this protein sequence reveals the following:

Possible site: 60

30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.2869(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

40 >GP:CAA66560 GB:X97918 gene 19.1 [Bacteriophage SPP1]
 Identities = 66/232 (28%), Positives = 106/232 (45%), Gaps = 12/232 (5%)

45 Query: 38 FRTLTVSGRDVVDLEHQTTSVLGRNGEYFHNATVEVRKLEIKAKISGKDNKS-MRLQYEK 96
 F V GR V +E ++ G +G ++ R+LE+ A + G ++ +R + E
 Sbjct: 24 FLVQEVRGRSVYSIEMGKRTIAGVDGGVITTESLPARELEVDAIVFGDGTETDLRRRIEY 83

Query: 97 LNKLI VSHNQVFLSFSDEPDRNYLGIFKSKDVP EEVSN E Q I I G L T F I C Y N P F K -----MS 151
 L N L + V ++FSDEP R Y G ++ +E + L F C +P K +
 Sbjct: 84 LNFLLRD TDVPITFSDEPSRTYYGRYEFATEGDEKGGFHKVTLN FYCQDPLKYGP EVTT 143

50 Query: 152 DVKTKKGTSIQNGGLFQTKPIITLNLSSPTKEIKLLHVESQKYIRLT----GTYYTDEIK 207
 DV T T ++N GL T P I S+ E ++ ++ ++ G T D +
 Sbjct: 144 DV-TTASTPVKNTGLAVTNPTIRCVFSTSATEYEMQLLDGSTVVKFLVKYGFNTGDTLV 202

55 Query: 208 IDMATGKITQNGRNILGDLDMINSRYFELLPGNNTLQCANAAITAEFREVYL 259
 ID +T NG++I+ L +I S + +L P NT A T F E +L
 Sbjct: 203 IDCHERSVTLNGQDIMPAL-LIQSDWIQLKPQVNTYLKATQPSTIVFTEKFL 253

-2759-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2686

A DNA sequence (GASx990) was identified in *S.pyogenes* <SEQ ID 7871> which encodes the amino acid sequence <SEQ ID 7872>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2861(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04681 GB:AP001510 unknown conserved protein in others

[Bacillus halodurans]

Identities = 116/449 (25%), Positives = 198/449 (43%), Gaps = 79/449 (17%)

Query: 2 IYLFDKLERLVATVG-TDDLSSWHFKVKNNDWDQASFEVPVDYDVEPFVYFGFFNYDPHQ 60

+++FD+ ++L+ T+ + L+ F+ + N F ++ E + + HQ

Sbjct: 4 LFIFDREDQLLTTLTESTGLVRALFREELNRVPNPQFAFTIEASSEEAKHV---IEEHQ 59

Query: 61 -----KEDVFKLFKVIDYNLEDSKFYKG-----LDKAESDLDTIAIKDKRFRQSSADA 109

KE +LF + + LED G + A +L I++ Q + +A

Sbjct: 60 VVFRDKEGDLRLFVIKE--LEDVDGLDGPQTATCEPAFMELAEHMIVEQSVVNQPAHEA 117

Query: 110 CIDGALEGTGYQVGKVEGITNVRTLSYYIISPRALIKIVEAFNCFNVRVYTF-INNKIT 168

++ AL+GT + G VE T + Y+S A+ I+ + +F TF N+IT

Sbjct: 118 -LNVALQGTRW-TGSVEVNLGNATEHFSYVSAIEAVWNILVTWGGDFKDVVTFNAENRIT 175

Query: 169 SRYIDLKKRFGKPTGKQFEHGNLLKVVVEESTDDIVTCLIGRGKGEEIQHEEAEPKDVE 228

S I + +R G GK+FE +N+ + + VT L GRG +Q E E +

Sbjct: 176 SHQIKIVQRRGVDRGKRFEIDHNI-EQIERTILSYPTALYGRGAS--LQGENGE----D 228

Query: 229 GHLPQEERRQGYGRRIEFTDVVWSVEKGDPIPKAGQNFVALDSAREEYGLSQNGELKHR 288

G L +F +V W G P+DKP GQ +V A ++YG NG+L HR

Sbjct: 229 GSL-----DFGEVEWRKSAGAPVDKPKGQLWVGDPALQKYGRKHNGQLLHR 275

Query: 289 WGVFVNEEIEDKTELLKATWEELQRLSIPIRIYKAEILDIGPETWKGDSVAIIYDEVKIA 348

G+F N IED ELL+ TWE+LQ+ S P Y+ + +++ +

Sbjct: 276 EGIFQNTNIEDPEELLEKTWEQLQKSSKPEVHYRLSVR-----LFEHIS-- 319

Query: 349 FETRVDEIDIDKLNFNRSVVTGLDYSVVQNR-----ESRSRKEAVQ-NMIDESLETITD 401

+ +LGD ++ +R E +SR A++ +++D + +

Sbjct: 320 -----GYEHEQASLGDTAIAIDRQFSRPIEQSRRIIAIEYDLVDIDGTGMVE 366

Query: 402 LGMTFQEFLLQGIKRIETGKKEMEDNWRK 430

+G L G+++R+E +E+E N K

Sbjct: 367 MGQFLS--LNGMDERLERIIEEIEKNQ GK 393

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2687

A DNA sequence (GASx991) was identified in *S.pyogenes* <SEQ ID 7873> which encodes the amino acid sequence <SEQ ID 7874>. Analysis of this protein sequence reveals the following:

-2760-

Possible site: 50

>>> Seems to have no N-terminal signal sequence

5 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2584(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

10 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA98101 GB:M19348 hyaluronidase [Streptococcus pyogenes phage
 H4489A]

Identities = 314/371 (84%), Positives = 338/371 (90%), Gaps = 1/371 (0%)

15

Query: 1 MAENIPLRVQFKRMKAAEWASSDVVLEGEIGFETDTGFAKFGDGQNTFSKLKYLTPGPKG 60
 M ENIPLRVQFKRM A EWA SDV+LLEGEIGFETDTGFAKFGDGQNTFSKLKYLTPGPKG
 Sbjct: 1 MTENIPLRVQFKRMSADEWARSVVILLEGEIGFETDTGFAKFGDGQNTFSKLKYLTPGPKG 60

20

Query: 61 PKGDTGLQGKTGGTSGRGPAGKPGTTDYDQLQNKPDLGAFQAQKEETNSKITKLESSKADK 120
 PKGDTGLQGKTGGT RGPAGKPGTTDYDQLQNKPDLGAFQAQKEETNSKITKLESSKADK
 Sbjct: 61 PKGDTGLQGKTGGTSGRGPAGKPGTTDYDQLQNKPDLGAFQAQKEETNSKITKLESSKADK 120

25

Query: 121 NAVYLKAESNAKLDEKLNKGGVMTGQLQFKPN-SGIKPSSSVGGAINIDMSKSEGAAMV 179
 +AVY KAES +LD+KL+L GG++TGQLQFKPN SGIKPSSSVGGAINIDMSKSEGAAMV
 Sbjct: 121 SAVYSKAESKIELDKKLSLTGGIVTGQLQFKPNKSGIKPSSSVGGAINIDMSKSEGAAMV 180

30

Query: 180 MYTNKDTTDDGPLMILRSNKDTFDQSVQFVDYKGTNAVNIVMRQPTTPNFSSALNITSAN 239
 MYTNKDTTDDGPLMILRS+KDTFDQS QFVDY G TNAVNIVMRQ+ PNFSSALNITSAN
 Sbjct: 181 MYTNKDTTDDGPLMILRSKDTFDQSAQFVDYSGKTNAVNIVMRQPSAPNFSSALNITSAN 240

35

Query: 240 EGGSAMQIRGVEKALGTLKITHENPSVDKEYDKNAAALSIDIVKKQKGGKGTAAQGIYIN 299
 EGGSAMQIRGVEKALGTLKITHENP+V+ +YD+NAAALSIDIVKKQKGGKGTAAQGIYIN
 Sbjct: 241 EGGSAMQIRGVEKALGTLKITHENPNVEAKYDENAAALSIDIVKKQKGGKGTAAQGIYIN 300

40

Query: 300 STSGTTGKLLRIRNLNDDKFPYVPGDGGFYAKETSQIDGNLKLKDPPIANDHAATKAYVDGE 359
 STSGT GK+LRIRN N+DKFYV PDGGF++ S + GNL +KDP + HAATK YVD +
 Sbjct: 301 STSGTAGKMLRIRNKNEKFPYVPGDGGFHSKANSTVAGNLTVKDPTSGKHAATKDYVDEK 360

Query: 360 VEKLLALLAAK 370
 + +LK L+ K
 Sbjct: 361 IAELKKLILKK 371

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 antigens for vaccines or diagnostics.

Example 2688

A DNA sequence (GASx993) was identified in *S.pyogenes* <SEQ ID 7875> which encodes the amino acid
 sequence <SEQ ID 7876>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1358(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2761-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2689

A DNA sequence (GASx995) was identified in *S.pyogenes* <SEQ ID 7877> which encodes the amino acid sequence <SEQ ID 7878>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0855 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC34418 GB:AF158600 gp149 [Streptococcus thermophilus
bacteriophage Sfil1]

Identities = 27/95 (28%), Positives = 50/95 (52%), Gaps = 2/95 (2%)

Query: 9 KYPQLDGTGAVASTHIIIAEDGAVIPQLIKQDLTSTNDTEIIKAALEEFKKSEYVEIAM 68
K + D +GA +T +I+ DGA +P + + ++TE++K ALE + + + A
Sbjct: 26 KSKEYDASGAAYATKVLKNRDGAYVPVFLPVEKIDLSNTELLKEALEVIYQENFPQRAE 85

Query: 69 GEAVQKVDLEKISQETAKTAKTAQTAAGLAKVSA 103
E ++D EKI + A + K +T A + + S+
Sbjct: 86 NEKFNELD--EKIKEYEALSKKATETIAKMEEASS 118

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2690

A DNA sequence (GASx996) was identified in *S.pyogenes* <SEQ ID 7879> which encodes the amino acid sequence <SEQ ID 7880>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -4.62 Transmembrane 9 - 25 (7 - 26)

----- Final Results -----

bacterial membrane --- Certainty=0.2848 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2691

A DNA sequence (GASx997) was identified in *S.pyogenes* <SEQ ID 7881> which encodes the amino acid sequence <SEQ ID 7882>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -3.66 Transmembrane 38 - 54 (35 - 55)

----- Final Results -----

bacterial membrane --- Certainty=0.2466(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2692

A DNA sequence (GASx998R) was identified in *S.pyogenes* <SEQ ID 7883> which encodes the amino acid sequence <SEQ ID 7884>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -9.87 Transmembrane 47 - 63 (41 - 72)

----- Final Results -----

bacterial membrane --- Certainty=0.4949(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2693

A DNA sequence (GASx999) was identified in *S.pyogenes* <SEQ ID 7885> which encodes the amino acid sequence <SEQ ID 7886>. Analysis of this protein sequence reveals the following:

Possible site: 24

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2763-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2694

A DNA sequence (GASx1001) was identified in *S.pyogenes* <SEQ ID 7887> which encodes the amino acid sequence <SEQ ID 7888>. Analysis of this protein sequence reveals the following:

Possible site: 22

```
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL      Likelihood = -10.51      Transmembrane      18 - 34 ( 16 - 34)
```

----- Final Results -----

```
      bacterial membrane --- Certainty=0.5203 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2695

A DNA sequence (GASx1002) was identified in *S.pyogenes* <SEQ ID 7889> which encodes the amino acid sequence <SEQ ID 7890>. Analysis of this protein sequence reveals the following:

Possible site: 32

```
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL      Likelihood = -3.61      Transmembrane      12 - 28 ( 11 - 33)
```

----- Final Results -----

```
      bacterial membrane --- Certainty=0.2444 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein is similar to AF186180 from *S.equi*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2696

A DNA sequence (GASx1003) was identified in *S.pyogenes* <SEQ ID 7891> which encodes the amino acid sequence <SEQ ID 7892>. Analysis of this protein sequence reveals the following:

Possible site: 32

```
>>> Seems to have a cleavable N-term signal seq.
```

----- Final Results -----

```
      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

-2764-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein is similar to SeeH from *S.equi*:

```

5  >GP:AAF72809 GB:AF186180 SeeH [Streptococcus equi]          Length = 236
    Identities = 233/236 (98%), Positives = 234/236 (98%)

Query: 1  MRYNCRYSHIDKKIYSMIICLSFLLYSNVVQANSYNTTNRHNLESYKHDSNLIADSIK 60
      Sbjct: 1  MRYNCRYSHIDKKIYSMIICLSFLLYSNVVQANSYNTTNRHNLESYKHDSNLIADSIK 60

10 Query: 61  NSPDIVTSHMLKYSVKDKNLSVFFEKDWISQEFKDKEVDIYALSAQEVCECPGKRYEAFG 120
      Sbjct: 61  NSPDIVTSHMLKYSVKDKNLSVFFEKDWISQEFKDKEVDIYALSAQE CECPGKRYEAFG 120

15 Query: 121  GITLTNSEKKEIKVPVNVWDKSKQPPMFITVKNPKVTAQEVDIKVRKLLIKKYDIYNNR 180
      Sbjct: 121  GITLTNSEKKEIKVPINVWDKSKQHPMPFITVKNPKVTAQEVDIKVRKLLIKKYDIYNNR 180

Query: 181  EQKYSKGTVTLDLNSGKDIVFDLYYFGNGDFNSMLKIYSNNERIDSTQFHVDSIS 236
      Sbjct: 181  EQKYSKGTVTLDLNSGKDIVFDLYYFGNGDFNSMLKIYSNNERIDSTQFHVDSIS 236
20

```

There is also homology to a *S.aureus* enterotoxin:

```

25 >GP:AAA19777 GB:U11702 enterotoxin H [Staphylococcus aureus]
    Identities = 70/215 (32%), Positives = 108/215 (49%), Gaps = 19/215 (8%)

Query: 27  SNVQANSYNTTNRHNLESYKHDSNLIADSI-KNSPDIVTSHMLKYSVKDKNLSVFFE 85
      +++ AN+Y N ++ K D E D I +N D +K++ D
Sbjct: 34  TDLALANAYGQYNHPFIKENIKSDEISGEKDLIFRNQGDSGNDRVKFATAD----- 85

30 Query: 86  KDWISQEFKDKEVDIYALSAQEVCECPGKRYEA--FGGITLTNSEK--KEIKVPVNVWDK 141
      ++Q+FK+K VDIY S CE + +GG TL NSEK +E + NVW
Sbjct: 86  ---LAQFKFNKNVDIYGASFYKCEKISENISECLYGGTTL-NSEKLAQERVIGANVWVD 141

35 Query: 142  SKQQPPMFITVKNPKVTAQEVDIKVRKLLIKKYDIYNNREQKYSKGTVTLDLNSGKDIVF 201
      Q+ I NK VT QE+DIK+RK+L KY IY ++ + SKG + D+ + +D F
Sbjct: 142  GIQKETELIRTNKKNVTLQELDIKIRKILSDKYKIY-YKDSEISKGLIEFDMKTTPRYSF 200

Query: 202  DLYYFGNGDFNSMLKIYSNNERIDSTQF-HVDVSI 235
      D+Y + + KIY +N+ + S H+DV++
40 Sbjct: 201  DIYDLKGENDYEDKIYEDNKTLSDDISHIDVNL 235

>GP:AAC26661 GB:AF064774 extracellular enterotoxin type I precursor
      [Staphylococcus aureus]
    Identities = 68/214 (31%), Positives = 109/214 (50%), Gaps = 27/214 (12%)
45

Query: 42  NLESY-KHDSNLIADSIKNSPDIVTSHMLKYSVKDKNLSVFFEKDWIS-QEFKDKEVD 99
      NL + Y KHD ++ + KN P ++ L++S +L + +W +FK K++D
Sbjct: 32  NLRNFYTKHDYIDLKGVTDKNLP---IANQLEFSTGTNDL-ISESNWDEISKFKGKKLD 87

50 Query: 100  IYALSAQEVCECPGKRYEAFGGITLTNSEKKEI-KVPVNVWDKSKQQPPMF--ITVKNPK 156
      I+ + C K +GG TL+ K+P+N+W K + I NK
Sbjct: 88  IFGIDYNGPC----KSKYMYGGATLSQOYLNSARKIPINLWVNGKHKTISTDKIATNKKL 143

55 Query: 157  VTAQEVDIKVRKLLIKKYDIYNNRE-----QKYSKGTVTLDLNSGKDIVFD 202
      VTAQE+D+K+R+ L ++Y+IY + ++ G V LN+ K +D
Sbjct: 144  VTAQEIADVKLRRYLQEEYNTIYGHNNRTGKGKEYGYKSKFYSGFNNGKVLPHLNNEKSFSYD 203

Query: 203  LYYFGNGDFNSMLKIYSNNERIDSTQFHVDSIS 236
      L+Y G+G S LKIY +N+ I+S +FH+DV IS
60 Sbjct: 204  LFYTGDLPLVPSFLKIYEDNKIIESEKFHLDVEIS 237

>GP:AAC28968 GB:U93688 enterotoxin [Staphylococcus aureus]
    Identities = 70/244 (28%), Positives = 127/244 (51%), Gaps = 27/244 (11%)
65

Query: 12  KKIYSMIICLSFLLYSNVVQANSYNTTNRHNLESYKHDSNLIADSIKNSPDIVTSHML 71

```

-2765-

KK+ S+++ ++ ++ A++ NL + Y + ++ +K++ D ++ L
 Sbjct: 2 KKLISILL-INIIILGVSNNASAQGDIGIDNLRNFYTK-KDFVDLKDVKDN-DTPIANQL 58
 Query: 72 KYSVKDKNLSVFFFEKDWIS-QEFKDKEVDIYALSAQEVCECPGKRYEAFGGITLTNSE-K 129
 ++S + +L + KD+ FK K++D++ +S C +Y +GG+T TN
 Sbjct: 59 QFSNESYDL-ISESKDFNKF'SNFKGKKLDVFGISYNGQCNT---KY-IYGGVTATNEYLD 113
 Query: 130 KEIKVPVNVW--DKSKQPPMFITVINKPKVTAQEVDIKVRKLLIKKYDIYNNREQK---- 183
 K +P+N+W K ++ NK VTAQE+D+K+RK L ++Y+IY + K
 Sbjct: 114 KSRNIPINIWINGNHKTISTNKVSTNKKLVTAQEI DVKLRKYLQEEYNIYGHNGTKKGEE 173
 Query: 184 -----YSGKGTVTLDLNSGKDIVFDLYYFG-NGDFNSMLKIYSNNERIDSTQFHVD 232
 ++ G VT LN+ +DL+Y G +G S LKIY +N+ ++S +FH+D
 Sbjct: 174 YGHKSKFYSGFNIGKVTFHLNNNDTFSYDLFYTGDDGLPKSFLKIYEDNKTVESEKPHLD 233
 Query: 233 VSIS 236
 V IS
 Sbjct: 234 VDIS 237

20 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2697

A DNA sequence (GASx1004R) was identified in *S.pyogenes* <SEQ ID 7893> which encodes the amino acid sequence <SEQ ID 7894>. Analysis of this protein sequence reveals the following:

25 Possible site: 29
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -2.18 Transmembrane 12 - 28 (12 - 28)
 30 ----- Final Results -----
 bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2698

40 A DNA sequence (GASx1009) was identified in *S.pyogenes* <SEQ ID 7895> which encodes the amino acid sequence <SEQ ID 7896>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence
 45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.6391(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2766-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2699

A DNA sequence (GASx1011) was identified in *S.pyogenes* <SEQ ID 7897> which encodes the amino acid sequence <SEQ ID 7898>. Analysis of this protein sequence reveals the following:

Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4528(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2700

A DNA sequence (GASx1024) was identified in *S.pyogenes* <SEQ ID 7899> which encodes the amino acid sequence <SEQ ID 7900>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2701

A DNA sequence (GASx1033) was identified in *S.pyogenes* <SEQ ID 7901> which encodes the amino acid sequence <SEQ ID 7902>. Analysis of this protein sequence reveals the following:

Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1652(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

-2767-

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2702

- 5 A DNA sequence (GASx1039) was identified in *S.pyogenes* <SEQ ID 7903> which encodes the amino acid sequence <SEQ ID 7904>. Analysis of this protein sequence reveals the following:

Possible site: 22

```

10 >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -1.06    Transmembrane    15 - 31 ( 15 - 31)

    ----- Final Results -----
    bacterial membrane --- Certainty=0.1426(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

- 20 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2703

A DNA sequence (GASx1058) was identified in *S.pyogenes* <SEQ ID 7905> which encodes the amino acid sequence <SEQ ID 7906>. Analysis of this protein sequence reveals the following:

```

25 Possible site: 60

    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.5484(Affirmative) < succ>
30    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

- 35 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2704

A DNA sequence (GASx1077) was identified in *S.pyogenes* <SEQ ID 7907> which encodes the amino acid sequence <SEQ ID 7908>. Analysis of this protein sequence reveals the following:

```

40 Possible site: 31

    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
45    bacterial cytoplasm --- Certainty=0.4848(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-2768-

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2705

A DNA sequence (GASx1080) was identified in *S.pyogenes* <SEQ ID 7909> which encodes the amino acid sequence <SEQ ID 7910>. Analysis of this protein sequence reveals the following:

Possible site: 40

```

10  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood =-12.42    Transmembrane  107 - 123 ( 93 - 133)
    INTEGRAL    Likelihood =-11.20    Transmembrane   20 - 36 ( 14 - 44)
    INTEGRAL    Likelihood = -8.39    Transmembrane  226 - 242 ( 218 - 246)
    INTEGRAL    Likelihood = -5.52    Transmembrane  129 - 145 ( 126 - 148)
15  INTEGRAL    Likelihood = -4.46    Transmembrane  160 - 176 ( 159 - 183)
    INTEGRAL    Likelihood = -1.44    Transmembrane   55 - 71 ( 55 - 72)

    ----- Final Results -----
    bacterial membrane --- Certainty=0.5967(Affirmative) < succ>
20  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2706

A DNA sequence (GASx1081) was identified in *S.pyogenes* <SEQ ID 7911> which encodes the amino acid sequence <SEQ ID 7912>. Analysis of this protein sequence reveals the following:

Possible site: 34

```

30  >>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood =-13.00    Transmembrane  103 - 119 ( 91 - 129)
    INTEGRAL    Likelihood =-11.46    Transmembrane  208 - 224 ( 203 - 230)
    INTEGRAL    Likelihood = -8.28    Transmembrane   54 - 70 ( 46 - 71)
35  INTEGRAL    Likelihood = -5.79    Transmembrane  160 - 176 ( 155 - 181)
    INTEGRAL    Likelihood = -4.25    Transmembrane  127 - 143 ( 125 - 149)

    ----- Final Results -----
    bacterial membrane --- Certainty=0.6201(Affirmative) < succ>
40  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2707

A DNA sequence (GASx1089) was identified in *S.pyogenes* <SEQ ID 7913> which encodes the amino acid sequence <SEQ ID 7914>. Analysis of this protein sequence reveals the following:

Possible site: 37

-2769-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2999(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

10 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2708

15 A DNA sequence (GASx1109) was identified in *S.pyogenes* <SEQ ID 7915> which encodes the amino acid sequence <SEQ ID 7916>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.1270(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2709

30 A DNA sequence (GASx1114R) was identified in *S.pyogenes* <SEQ ID 7917> which encodes the amino acid sequence <SEQ ID 7918>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

35

----- Final Results -----

 bacterial cytoplasm --- Certainty=0.4021(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2770-

Example 2710

A DNA sequence (GASx1149) was identified in *S.pyogenes* <SEQ ID 7919> which encodes the amino acid sequence <SEQ ID 7920>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -1.70 Transmembrane 12 - 28 (12 - 29)

----- Final Results -----

bacterial membrane --- Certainty=0.1680 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2711

A DNA sequence (GASx1150) was identified in *S.pyogenes* <SEQ ID 7921> which encodes the amino acid sequence <SEQ ID 7922>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2712

A DNA sequence (GASx1160) was identified in *S.pyogenes* <SEQ ID 7923> which encodes the amino acid sequence <SEQ ID 7924>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -3.19 Transmembrane 15 - 31 (15 - 31)

----- Final Results -----

bacterial membrane --- Certainty=0.2275 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2771-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2713

A DNA sequence (GASx1167) was identified in *S.pyogenes* <SEQ ID 7925> which encodes the amino acid sequence <SEQ ID 7926>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1404(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB99233 GB:U67563 oxaloacetate decarboxylase alpha chain (oadA)
[Methanococcus jannaschii]
Identities = 250/453 (55%), Positives = 325/453 (71%), Gaps = 7/453 (1%)

Query: 13 VAITETVLRDGHQSLMATRLSIEDMLPVLITLIDKIGYYSLCWWGATFDACIRFLNEDPW 72
V I +T RD QSL+ATR+ EDMLP+ +D++G+YS+E WGGATFDACIR+LNEDPW
Sbjct: 2 VKIVDTTFRDAQQSLIATRMRTEDMLPIAEKMDVEVGFYSMEVWGGATFDACIRYLNEDPW 61

Query: 73 ERLRTLKKGLPENTRLQMLLRGQNLGYPHYADDIVDKFISLSAQNGIDVFRIFDALNDPR 132
ERLR LKK + NT LQMLLRGQNL+GYRHY DDIV+KF+ + +NGID+FRIFDALND R
Sbjct: 62 ERLRALKKRIQNTPLQMLLRGQNLVGYRHYPDIVKFKVIAHENGIDIFRIFDALNDVR 121

Query: 133 NIQQALRAVKKTGKEAQLCIAYTTSFVHTLNYYLSLVKELVEMGADSICIKDMAGILTPK 192
N++ A++ KK G E Q I YT SPVHT++ Y+ L K+L EMG DSICIKDMAG+LTP
Sbjct: 122 NMETAIKTAKKVGAEVQGAICYTISFVHTIDQYVELAKKLEEMGCDSICIKDMAGLLTPY 181

Query: 193 AAKELVSGIKAMTNLPLIVHTHATSGISQMTYLA AVEAGADRIDTALSPFSEGTSQPATE 252
ELV +K +LP+ VH+H TSG++ MYL +EAGAD +D A+SPF+ GTSQP TE
Sbjct: 182 EGYELVKRLKEEISLPIDVHSHCTSG LAPMTYLVKIEAGADMVDC AISPFGMTSQPPTE 241

Query: 253 SMYLALKEASYDITLDETLLQEAANHLRQARQKYLADGILDPSLLFPDPRTLQYQVPGGM 312
S+ +ALK YD LD LL + ++ + R+KY + P D R L YQVPGGM
Sbjct: 242 SIVVALKGTKYDTGLDLKLLNEIRDYFMKVREKYKM--LFSPISQIVDARVLVYQVPGGM 299

Query: 313 LSNMLSQKQANAEKLEEVLAEPVRVRKDLGYPLVTPLSQMVGTQAAMNVILGKPYQM 372
LSN++SQLK+ A K EEVL E+PRVRKDLGYPLVTP SQ+VGTQA +NV+ + Y++
Sbjct: 300 LSNLVSQKQAGALDKFEVLQEI PRVRKDLGYPLVTP TQSIVGTQAVLNVLTEERYKI 359

Query: 373 VSKEIKQYLAGDYGKTPAPVNEDLKRSQI--GSAPVT TNRPADQLSPEFEVLK--AEVAD 428
++ E+ Y+ G YGK PAP+N +L + + G P+T RPAD L PE+E +K AE
Sbjct: 360 ITNEVVNYVKGFYKPPAPINPELLKRVLDGEKPITC--RPADLLPPEWEKVKKEAEKEG 418

Query: 429 LAQTDDEVLTYALFSPVAKPFLITTKYQTD DVIK 461
+ + +ED+LTYAL+P +A FL + + + + K
Sbjct: 419 IVKKEEDILTALYPQIAVKFLRGELKAEPIPK 451

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2714

A DNA sequence (GASx1168) was identified in *S.pyogenes* <SEQ ID 7927> which encodes the amino acid sequence <SEQ ID 7928>. Analysis of this protein sequence reveals the following:

-2772-

Possible site: 38

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -7.11 Transmembrane 16 - 32 (2 - 34)

----- Final Results -----

bacterial membrane --- Certainty=0.3845(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2715

A DNA sequence (GASx1170) was identified in *S.pyogenes* <SEQ ID 7929> which encodes the amino acid sequence <SEQ ID 7930>. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -7.06 Transmembrane 211 - 227 (208 - 238)

INTEGRAL Likelihood = -5.84 Transmembrane 117 - 133 (110 - 136)

INTEGRAL Likelihood = -5.36 Transmembrane 256 - 272 (253 - 274)

INTEGRAL Likelihood = -4.67 Transmembrane 44 - 60 (41 - 64)

INTEGRAL Likelihood = -4.19 Transmembrane 287 - 303 (287 - 306)

INTEGRAL Likelihood = -3.77 Transmembrane 358 - 374 (357 - 375)

INTEGRAL Likelihood = -2.18 Transmembrane 20 - 36 (16 - 38)

INTEGRAL Likelihood = -0.85 Transmembrane 90 - 106 (90 - 106)

INTEGRAL Likelihood = -0.53 Transmembrane 165 - 181 (164 - 181)

----- Final Results -----

bacterial membrane --- Certainty=0.3824(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA05140 GB:AJ002015 methylmalonyl-CoA decarboxylase,

beta-subunit [Propionigenium modestum]

Identities = 231/395 (58%), Positives = 293/395 (73%), Gaps = 19/395 (4%)

Query: 1 MLDVNLQMVQSSGLAHLTVNNLMICLASFFLYLGIKKEYEPYLMVPIAFGILLVNLPM 60

ML + S+G L + ++IM+ +A FLYL I KE+EP L+VPI+FGILL NLP A

Sbjct: 1 MLQAILDFYHSTGFYGLNMGSIIMMLVACVFLYLAIKEFEPLLLVPISEFGILLTNLPFA 60

Query: 61 GLMDHP-----ANG-----NPGGLLYLYLKGTSLGIYPPLIFLCLGASTDFG 102

G+M P A+G PGGLLYYL++G LGI+PPLIFL +GA TDFG

Sbjct: 61 GMMAEPLLEVHEKLSASGAHLYTAHTAEPGGLLYYLFQGDHLGIFPPLIFLGVGAMTDFG 120

Query: 103 PLIANPKTILLGGAQVGIFLAFFLAIMLGM-TPQEASVGIIGGADGPTAIYVTTKLAP 161

PLI+NPK++LLG AAQ GIF+ FF AI G+ T QEAS+GIIGGADGPTAI++++KLAP

Sbjct: 121 PLISNPKSLLLGAAQGFIVTFFGAIASGLFTAQEASIGIIGGADGPTAIFLSSKLAP 180

Query: 162 DLLSTIALAAYSIMALVPPIQPPIIKLLTTKAERQVKMTQARTVSQKEKIIFPIMVTIFV 221

L+ IA+AAYSIMALVPPIQPPI+ LT++ ER++KM+Q R VS++EKIIFPI+VTI V

Sbjct: 181 HLMGPPIAAYAAYSIMALVPPIQPPIIMTALTSETERKIKMSQLRLVSKREKIIFPIVVTILV 240

Query: 222 SLLVPSATTLVGCLMLGNLVREIKIVPKIVENLQQVVMFCITIIILGLTVGAKANGDLFLS 281

-2773-

SL+VP A TLVG LMLGNL RE +V ++ + + ++ ITI LG+TVGA A + FL
 Sbjct: 241 SLIVPPAATLVGMLMLGNLFRECGVVGRLED TAKNALINIITIFLGVTVGATATAEAFK 300

Query: 282 ATTLKIIALGLIAFAAGTAGGVLGKVMYYLSGNKVNPMIGAAGVSAVPMARVVQKIGQ 341
 TL I+ LG++AF GT GVL+ K M LS +NP++G+AGVSAVPMARV Q +GQ
 Sbjct: 301 VETLAILGLGIVAFGIGTSGVLLAKFMNKLKEPINPLLGSAGVSAVPMARVSQVVQ 360

Query: 342 EEDPSNFLMHAMGPNVAGVIGSAIASGALLAF 376
 + DP+NFLLMHAMGPNVAGVIGSA+++G LL+ FG
 Sbjct: 361 KADPTNFLMHAMGPNVAGVIGSAVSAGVLLSLFG 395

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2716

A DNA sequence (GASx1171R) was identified in *S.pyogenes* <SEQ ID 7931> which encodes the amino acid sequence <SEQ ID 7932>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0851(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF93965 GB:AE004165 citG protein [Vibrio cholerae]
 Identities = 100/287 (34%), Positives = 154/287 (52%), Gaps = 12/287 (4%)

Query: 9 ISQLALKALLYEVSLSPKPGVLDRFDNGAHDDMSFITFIDSMIALSPFFQAYIETGFAYA 68
 + LA A++ EV L+PKPGLVD +NGAH DM TFI S A++P+ +++ G+ A
 Sbjct: 32 VGHAYHAMMLEVHLTPKPGGLVD TANNGAHRDMDLNTFIASAEAIAPYLHSFVSAGWESA 91

Query: 69 KEEPLLLFNRLRQLGQKAEETMFCATQGINTHKGLNFSMALLIGATGAYLARTPHLMTDL 128
 L + LR +G +AE+ MF ATQG+NTHKG+ F + L+ G+ G A
 Sbjct: 92 GNPAQLLSALRPIGIEAEQAMFAATQGVNTHKGMIFILGLICGSVGWLKANQ----- 144

Query: 129 GRFSKEDTLAICRLVKPMTAHLIQTDLGHLNTHKKEFTYGEQLFVTYGIKGPRGEASEGFT 188
 K D I ++ L+ +L + T GE+++ YG+ G RGEA+ G
 Sbjct: 145 ---LKIDAQHIGETIRQACQFLVIDELKAKRDCEPETAGERIYRQYGLTGARGEAAASGLA 201

Query: 189 TLTDHALPYFRQMISON-DPETSQRLRLVYLSIVEDGNLIHRGGIEAWKGVKAD-MRL 246
 + HALP ++ +++ E + L+ LM+ D NL+ RGG+ V+ +LL
 Sbjct: 202 MVMQHALPAYQACLTGASTEQALWHTLLVLMANNNDNLVSRGGLAGLHFVQEQAQQL 261

Query: 247 LQQDLSTTDRLALSSYNQCLINQHLSPGGAADLLALTFYFAPLEKL 293
 + ++ AL++ + LI +HLSPGG+ADLLA T+ L +L
 Sbjct: 262 AKGGFLYQIEQALTALDSVLIEKHLSPGGSADLLAATWLIYELVQL 308

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2717

A DNA sequence (GASx1172R) was identified in *S.pyogenes* <SEQ ID 7933> which encodes the amino acid sequence <SEQ ID 7934>. Analysis of this protein sequence reveals the following:

Possible site: 23

-2774-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2501(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

10 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12389 GB:Z99107 similar to transcriptional regulator (GntR
 family) [Bacillus subtilis]
 Identities = 60/205 (29%), Positives = 99/205 (48%), Gaps = 3/205 (1%)

15 Query: 19 PLKIAFYNALKKTII LRQIPVGS RINEKEFSIALNISRTPIRYALG LLLSEEHLVEHIPKK 78
 P + FYN LKK I G RINE + + + +SR+PIR A+ LL ++ L++ +
 Sbjct: 11 PYYLQFYNQLKKMIFNGTFKPGERINETQLAKSFGVSRSPIREAMRLLEKDGLLKADDRN 70

20 Query: 79 GIIVKGVSIKDACEIFEIRKALETLATVQAMHLMTEEDFKVMHNLEDCETFI--AEDDT 136
 G + ++ KD EI++IR LE LA + EE+ ++ LE+ E I +DT
 Sbjct: 71 GFSITSLTAKDVDEIYKIRIPLQLAVELVIDEADDEEELTILEKQLEETEKAHNGTDT 130

25 Query: 137 NRILDNFNAFNNLIYSYSQMVRLKEIVTELQAYLVYFRKISSISVERRKRALSEHWMIIYR 196
 I N F+ L+ +S LK ++ + + + R ++ + R + L EH I+
 Sbjct: 131 EIIRLN-QKFHELLVDFSHNRHLKNLLEHVNDLIHFCRILNYTGDHRAETILREHRRIFE 189

30 Query: 197 GMKNKDHEQITLITHEHLNSSLFI 221
 +K K+ E H N E +
 Sbjct: 190 EVKKKNKEAAKQHVLAHFNHDCEHL 214

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2718

A DNA sequence (GASx1173R) was identified in *S.pyogenes* <SEQ ID 7935> which encodes the amino
 35 acid sequence <SEQ ID 7936>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have a cleavable N-term signal seq.

40 INTEGRAL Likelihood = -10.99 Transmembrane 450 - 466 (445 - 473)
 INTEGRAL Likelihood = -9.61 Transmembrane 33 - 49 (30 - 55)
 INTEGRAL Likelihood = -8.55 Transmembrane 326 - 342 (321 - 346)
 INTEGRAL Likelihood = -7.01 Transmembrane 288 - 304 (286 - 311)
 INTEGRAL Likelihood = -6.79 Transmembrane 95 - 111 (88 - 114)
 INTEGRAL Likelihood = -4.99 Transmembrane 265 - 281 (264 - 285)
 45 INTEGRAL Likelihood = -4.62 Transmembrane 208 - 224 (204 - 228)
 INTEGRAL Likelihood = -3.13 Transmembrane 126 - 142 (126 - 145)
 INTEGRAL Likelihood = -2.81 Transmembrane 366 - 382 (365 - 383)
 INTEGRAL Likelihood = -2.34 Transmembrane 419 - 435 (417 - 438)

50 ----- Final Results -----

 bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 A related sequence was also identified in GAS <SEQ ID 9169> which encodes the amino acid sequence
 <SEQ ID 9170>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 39

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -10.99 Transmembrane 443 - 459 (438 - 466)

-2775-

5
 INTEGRAL Likelihood = -8.55 Transmembrane 319 - 335 (314 - 339)
 INTEGRAL Likelihood = -7.01 Transmembrane 281 - 297 (279 - 304)
 INTEGRAL Likelihood = -6.79 Transmembrane 88 - 104 (81 - 107)
 INTEGRAL Likelihood = -4.99 Transmembrane 258 - 274 (257 - 278)
 INTEGRAL Likelihood = -4.62 Transmembrane 201 - 217 (197 - 221)
 INTEGRAL Likelihood = -3.13 Transmembrane 119 - 135 (119 - 138)
 INTEGRAL Likelihood = -2.81 Transmembrane 359 - 375 (358 - 376)
 INTEGRAL Likelihood = -2.34 Transmembrane 412 - 428 (410 - 431)

10
 ----- Final Results -----
 bacterial membrane --- Certainty=0.539(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG08853 GB:AE004959 probable citrate transporter [Pseudomonas aeruginosa]
 Identities = 199/468 (42%), Positives = 296/468 (62%), Gaps = 41/468 (8%)

20
 Query: 9 LLTMLAYAMIIVFMYVVMKKMTPTALVMIPLIMTIAVILTGSADFNADAKFVAFVGDG 68
 +LT+LA+AM+ FM+++M K+++ AL+++P +AF G
 Sbjct: 1 MLTLAFAMVATFMFLIMTKRLSALIALILVP-----IAFALIG 39

25
 Query: 69 GIAKDLTAIGPMVMYGINNTAKTGIMLLFAILFFSVMLDAGLFDPITEKMIRFAKGDPMK 128
 G A L GPM++ GI A TG+ML+FAIL+F++M+D+GLFDP K++R KGDP+K
 Sbjct: 40 GFAAGL---GPMMLDGIRTLAPTGVMLMFAILYFAIMIDSGLFDPFAVRKILRLVKGDPK 96

30
 Query: 129 VLIATAVVAASVSLNGDGTITLICCFAFLPIYKKLDMKIMNLGVLIILQNTIMNLLPWG 188
 V + TA +A VSL+GDG+TT +IC +A LP+Y +L M + + LI+L + ++N+ PWG
 Sbjct: 97 VSLGTAALAMIVSLDGDGSTTYMICVAAVLPLYSRLGMSPLVMACLIMLSSGVNLNMTDPWG 156

35
 Query: 189 GPTARAMSVLGVGP-EILGYLAPGMILSLL--YVICWVAPSMGRKERARLGVIDL--SEE 243
 GPTARA S L V P +I + P MI LL + I W+ G++ERARLG + L E
 Sbjct: 157 GPTARAASALHVDPADIFVPMIPAMIAGLLAIFATAWI---YGKREERARLGELHLPDHE 213

40
 Query: 244 DMRQLTDITDPDTLFIIRPKNFVFNAILTIGLITWLVAGSFNKSAMAPLLLFAVGTCIA 303
 D+ +++ P+ RRPK FNAILT+ L+ L+AG + M L + A G IA
 Sbjct: 214 DLAEISVSQYPEA---RRPKLLWFNAILTVVLMATLIAGL----LPMPVLFMIAFG--IA 264

45
 Query: 304 LMVNPVVKDQSKRIGDNAGDAVQVVILVFAAGIFMGLFQSGMASALAQSFIATIPKQL 363
 ++VNYP +++Q KRIG +A + + VV L+FAAG+F G+ G+GM A+++S +IP L
 Sbjct: 265 MIVNYPICIQEQKKRIGAHENILAVVSLIFAAGVFTGILSGTGMVDAMSKSLAVIPPAL 324

50
 Query: 364 AGFWGLVIALVSAPGTFFFISNDGFYYGILPVLAEAGA EYGFSNMAMALASLMGQAFHLLS 423
 + + ALVS P TFF+SND FYYG+LP+L +A AEYG + + MA AS++GQ HLLS
 Sbjct: 325 GPYLATITALVSMPTFFMSNDAFYYGVLPILTQAAAEYGITPVEMARASIVGQPVHLLS 384

Query: 424 PLVAFIYLLRLTGLDMGEWQKEAAKYALIIFVIFVVTIIMGQMPLY 471
 PLV YLL+ L +D G+ Q+ K+A+++ + + + +G PL+
 Sbjct: 385 PLVPSTYLLVGLAKIDFGDHRFTLKWAVLVCLAILAMALLGLFPLF 432

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2719

55 A DNA sequence (GASx1174) was identified in *S.pyogenes* <SEQ ID 7937> which encodes the amino acid sequence <SEQ ID 7938>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have no N-terminal signal sequence

60
 ----- Final Results -----

-2776-

```

bacterial cytoplasm --- Certainty=0.3948(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

5 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2720

10 A DNA sequence (GASx1175) was identified in *S.pyogenes* <SEQ ID 7939> which encodes the amino acid sequence <SEQ ID 7940>. Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.3519(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

20 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 2721

A DNA sequence (GASx1177) was identified in *S.pyogenes* <SEQ ID 7941> which encodes the amino acid sequence <SEQ ID 7942>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have an uncleavable N-term signal seq

```

INTEGRAL Likelihood = -9.24 Transmembrane 115 - 131 ( 105 - 137)
INTEGRAL Likelihood = -8.92 Transmembrane 208 - 224 ( 204 - 238)
INTEGRAL Likelihood = -7.80 Transmembrane 282 - 298 ( 273 - 303)
INTEGRAL Likelihood = -4.94 Transmembrane 85 - 101 ( 75 - 102)
INTEGRAL Likelihood = -4.04 Transmembrane 10 - 26 ( 3 - 32)
INTEGRAL Likelihood = -3.61 Transmembrane 255 - 271 ( 253 - 271)

```

----- Final Results -----

```

bacterial membrane --- Certainty=0.4694(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAB89172 GB:AE000960 oxaloacetate decarboxylase, sodium ion pump
subunit (oadB) [Archaeoglobus fulgidus]
Identities = 190/354 (53%), Positives = 255/354 (71%), Gaps = 8/354 (2%)

Query: 16 IVMVIGALLMYLGIKKEYEPTLLVPMGLGTILVNFPGSGVLTQVVNGVEQEGVFEALFN 75
+VM+ +G LL+YLG I K+ EP LLVP+G+G ILVN PG G+ E+ +F+
Sbjct: 5 IVMIGVGLLLVYLGIIVKKMEPLLLVPIGIGAILVNIPGGGL-----AEEGSIFDLFLK 57

```

-2777-

5
 Query: 76 FGIGTELPILLIFIGIGAMIDFGPLLQNPFMLLFGDAAQGIFFVVVVAVLAGFDIKEAA 135
 + I TE+ PLLIF+G+GA+ DF PLL NP L G AAQ GIF ++ A+ GF +EAA
 Sbjct: 58 YLIHTEIVPLLIFLGLGALTDFSPLLANPKTFLLGAAAQIGIFAALIAALFLGFTPQEAA 117

10
 Query: 136 SIGIIGAADGPTSIFVANQLAKDLGPITVAAYSIMALVPPIQPFAIKLVTTKKERRIRM 195
 SIGIIG ADGPT+I+ LA LL VAAYSIM+LVPIIQP IK +T+ +ER+I+M
 Sbjct: 118 SIGIIGGADGPTTIYTTTTILAPHLAATAVAAYSIMSLVPPIQPPIIKALTSSRERKIKM 177

15
 Query: 256 VNIISILLGLTISIKMQADLFNLVQTLIIIVFGLLAFIMDSIGGVMFVAKFLNLFKEKIN 315
 +NI++I+LGL++ M+A+ FL +TLL++ G++AF + GG+V+ AK +NLF KEKIN
 Sbjct: 237 MNIMTIILGLSVGSTMRAESFLTQKTLVLALGVVAFAAATAGGVLLAKVMNLFKEKIN 296

20
 Query: 316 PMIGAAGISAFPMSSRVIQKMATDEDPQNFILMYAVGANVSGQIASVIAGGLLL 369
 PMIGAAG+SA PMS+RV+Q++A +EDP N ILM+A+G NV+G I S +A G+L+
 Sbjct: 297 PMIGAAGVSAVPM SARVVQRLAIEDPHNHILMHAMGPNVAGVIGSAVAAGVLI 350

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2722

25 A DNA sequence (GASx1178) was identified in *S.pyogenes* <SEQ ID 7943> which encodes the amino acid sequence <SEQ ID 7944>. Analysis of this protein sequence reveals the following:

Possible site: 16

30 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.50 Transmembrane 21 - 37 (8 - 43)

----- Final Results -----
 bacterial membrane --- Certainty=0.4800(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

40 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2723

A DNA sequence (GASx1179) was identified in *S.pyogenes* <SEQ ID 7945> which encodes the amino acid sequence <SEQ ID 7946>. Analysis of this protein sequence reveals the following:

45 Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1906(Affirmative) < succ>
 50 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

-2778-

>GP:AAF93961 GB:AE004165 citrate lyase, gamma subunit [Vibrio cholerae]
Identities = 46/97 (47%), Positives = 64/97 (65%)

Query: 1 MDIKQTAVAGSLESSDLMTTVSPNDEQTTITLDSSEKQFGNHIRQLIHQTLVNLKVTA 60
M I A AG+LESSDL + + PN++ I + LDS+VE+QFG+ IRQ++ TL ++V
Sbjct: 1 MKIAHPAFAGTLESSDLQVRIEPPNDGGIELVLDSTVEQQFGHAIRQVVLHTLDAMQVRD 60

Query: 61 AKVEAVDKGALDCTIQARTIAAVHRAAGIDQYDWKEI 97
A V DKGALDC I+AR AAV RA + +W ++

Sbjct: 61 ALVTIEDKGALDCVIRARVQAAMRACDVQNIWSQL 97

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2724

A DNA sequence (GASx1181) was identified in *S.pyogenes* <SEQ ID 7947> which encodes the amino acid sequence <SEQ ID 7948>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.65 Transmembrane 74 - 90 (74 - 90)

----- Final Results -----

bacterial membrane --- Certainty=0.1659(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA71632 GB:Y10621 CILB, citryl-CoA lyase beta subunit
[Leuconostoc mesenteroides]

Identities = 187/293 (63%), Positives = 237/293 (80%), Gaps = 1/293 (0%)

Query: 2 ERLRRTMMFVPGANAAMLRDAPLFGADSIMFDLEDSVSLKEKDTSRALVHFALKTFDYSS 61

ERLRRTMMFVPG N AM++DA +FGADSIMFDLED+VSL EKD++R LV+ AL+T DY S

Sbjct: 4 ERLRRTMMFVPGNNPAMVKDAGIFGADSIMFDLEDAVSLAEKDSARYLVYEALQTVDYGS 63

Query: 62 VETVVRVNGLDS-CGALDIEAVVLAVGNVIRLPKTETAQDIIDVEAVIERVERENSIEVG 120

E VVR+NGLD+ DI+A+V AG++VIRLPK ETA + ++E++I E+E VG

Sbjct: 64 SELVVRINGLDTPFYKNDIKAMVKAGIDVIRLPKVETAAMMHELESITDAEKEFGRPVG 123

Query: 121 RTRMMAAIESAEGVLNAREIAKASKRLIGIALGAEDYVTNMKTRRYPDGQELFFARSMIL 180

T MMAAIESA GV+NA EIA AS R+IGIAL AEDY T+MKT RYPDGQEL +AR++IL

Sbjct: 124 TTHMMAAIESALGVVNAVEIANASDRMIGIALSAEDYTTDMKTHRYPDGQELLYARNVIL 183

Query: 181 HAARAAGIAAIDTVSDVNNTTEGFQNEVRMIKQLGFDGKSVINRQIPLVNEIYTPTKKE 240

HAARAAGIAA DTV++++N+ EGF E ++I QLGF DGKS+INPRQI +VN++Y PT+KE

Sbjct: 184 HAARAAGIAAFDTVFTNLNDEEGFYRETQLIHQLGFDGKSLINRQIEMVNKVYAPTEKE 243

Query: 241 IDHAKQVIWAIREAESKSGVISLNGKMVDKPIVERAERVIALATAAGVLSEE 293

I++A+ VI AI EA+ KSGSVIS+NG+MVD+P+V RA+RV+ LA A ++ E

Sbjct: 244 INNAQNVIAAIEEAKQKSGVISMNGQMVDPRPVVILRAQRMKLANANHLVDSE 296

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2779-

Example 2725

A DNA sequence (GASx1182) was identified in *S.pyogenes* <SEQ ID 7949> which encodes the amino acid sequence <SEQ ID 7950>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3554 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAA71633 GB:Y10621 CILA, citrate CoA-transferase alpha subunit
[Leuconostoc mesenteroides]
Identities = 294/511 (57%), Positives = 378/511 (73%), Gaps = 7/511 (1%)

Query: 4   NKLGRDIPQPYADQY--GVFEGELANIKQYDESSRIKPVKPGDSKLLGSVREAIKTGL 61
          NK+  D+P   +Q   VFE      +      +++   G+SK+  S+  +  +  T L
Sbjct: 3   NKNVIDVPDAILEQLDDSVFESTNYGNPEIQRVGPKVRATT-GESKVVQSSIDDLVLSNT-L 60

Query: 62   TDGMTISFHHHFFREGDFIMNMVLEELIAKMGIKNLSIAPSSIANV-HEPLIDHIKNGVVTN 120
          DGMTISFHHHFFREGDF+ N V+ +I  MG +NL++APSS+ NV ++ +I+ IK GVVTN
Sbjct: 61   KDGMTISFHHHFFREGDFVFNKVMRKIIDMGYQNLTLAPSSLTNVMNDIVIEAIKKGVVTN 120

Query: 121  ITSSGLRDKVGAAISEGLMENPVVIRSHGGRARAIASGDIHIDVAFLGAPSSDAYGNVNG 180
          ITSSG+R  +G A+S G+++NPV+ RSHG RARAI SG+I IDVAFLG P+SD  GN NG
Sbjct: 121  ITSSGMRGTGLGDAVSHGILKNPVIFRSHGARARAIESGEIKIDVAFLGVPNSDEMGNANG 180

Query: 181  TKGKATCGSLGYAMIDAKYADQVILTDNLVYPNPNTPISSIPQTDVDYVVTVDAIGDPQGI 240
          G A  GSLGYA+IDA+YAD++V++TD ++PYPNTP SI QT VDYVV VD +GDP  I
Sbjct: 181  MNGDAAFGSLGYALIDAQYADKLVLITDTIMPYPNTPASIKQTQVDYVVKVDKVGDPDKI 240

Query: 241  AKGATRFTKPNKELLIAEYAAKVITNSPYFKEGFSFQTGTGGASLAVTRFMREAMIKENI 300
          GATRFTK+PKEL IA+  VI NS YFK  FSFQTG+GGA+LAVTRF+REAM+ +NI
Sbjct: 241  GSGATRFTKDPKELKIAKTVDNDVIVNSKYFKNDFSQTGSGGAALAVTRFLREAMMAQNI 300

Query: 301  KASFALGGITNAMVELLEELVEKILDVQDFDHPSAVSLGKHAHYEIDANMYASPLSKG 360
          ASFALGGIT  V+LL E LV +++DVQDFD  +A S+      EIDA+ YA P +KG
Sbjct: 301  MASFALGGITKPTVDLLNEGLVNRVMDVQDFDKGAASSMKLSPNQQEIDASWYADPANKG 360

Query: 361  AVINQLDTCILSALEVDTNFNVNVMTGSDGVIRGASGGHCDTAAFAAKMSLVISPLIRGRI 420
          A++++LD  ILSALEVDTNFNVNVM+GSDGVIRGA GGH D A AK++++  PL+RGRI
Sbjct: 361  AMVDKLDVAILSALEVDTNFNVNVMGSDGVIRGAIGGHQDAA-TAKLTIISVPLVRGRI 419

Query: 421  PTFVDEVNTVITPGTSVDVIVTEVGIAINPNRQDLVDHFKSL-NVPQFSIEELKEKAYAI 479
          T V +VNTVITPG S+DV+VTEVGIAINP R DLV+  K +  +P +SIEEL++KA  I
Sbjct: 420  ATIVPKVNTVITPGDSIDVVVTEVGIAINPKRTDLVEQLKQVPLPIYSIEELQQAIEKI 479

Query: 480  VGTPERIQYGDKVVALIEYRDGSLMDVVYINV 510
          VG P  +++ D+VVA+ EYRDGS++D++  V
Sbjct: 480  VGQAPPLKFTDRVVAEYRDGSEVIDIIEV 510

```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2726

A DNA sequence (GASx1183) was identified in *S.pyogenes* <SEQ ID 7951> which encodes the amino acid sequence <SEQ ID 7952>. Analysis of this protein sequence reveals the following:

-2780-

Possible site: 13

>>> Seems to have a cleavable N-term signal seq.

5 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA71634 GB:Y10621 CILG, hypothetical protein [Leuconostoc
 mesenteroides]
 Identities = 65/176 (36%), Positives = 97/176 (54%), Gaps = 3/176 (1%)
 15 Query: 21 DTYFSGEAIQLSDMLRAREERALRQLHLLKEYPEGSLLSVTMNIPGIKTSPKLLLEAFDI 80
 D + GE + L +L RE R Q L+ +P + SV +N+PGPIKTSPKL F I
 Sbjct: 2 DYFEGGERLNLMOVLNREWREKYQQLMASFPTAVITSVKLNLPGIKTSPKLQSVFQI 61
 20 Query: 81 VIKAIQTALADDKICYQLRL-PTTGYEYLLITSLPSRDLKLMIALETELPIGRIMDL 139
 +I + D +I + + TG + + +TS + +K MI E +GRL+DL
 Sbjct: 62 IINDLNPVFKDLQIIKEASFVDQITGPDIFVTSGCLKLVKQIMITFEESHLLGRLLDL 121
 25 Query: 140 VLVLQNDLPHSISRVTLVGGSPRQCFICSKEAKVCGRLRKHSVEEMQTAISKLLHSF 195
 V+ D +SR LG +PR+C +C K+AK C + HS+ E + I+K+LH+F
 Sbjct: 122 VMCQNAD--KQLSREELGFAPRKCLLCGKDAKTCIKEGNHSLAEGYSQINKMLHNF 175

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 2727

A DNA sequence (GASx1184) was identified in *S.pyogenes* <SEQ ID 7953> which encodes the amino acid sequence <SEQ ID 7954>. Analysis of this protein sequence reveals the following:

Possible site: 58

35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3730(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB99233 GB:U67563 oxaloacetate decarboxylase alpha chain (oadA)
 [Methanococcus jannaschii]
 45 Identities = 245/441 (55%), Positives = 336/441 (75%), Gaps = 5/441 (1%)
 Query: 10 IRITETVLRDQQSQIATRM TTKEMIPILETLDNAGYHALEMWGGATFDSCLRFLNEDPW 69
 ++I +T RD QQS IATRM T++M+PI E +D G++++E+WGGATFD+C+R+LNEDPW
 50 Sbjct: 2 VKIVDTTFRDAQQSLIATRMRTEDMLPIAEKMDVEGVFYSMEVWGGATFDACIRYLNEDPW 61
 Query: 70 ERLRAIRKAVKTKLQMLLRGQNLGYRNYADDVVSFIQKSIENGIDIVRIFDALNDPR 129
 ERLRA++K ++ T LQMLLRGQNL+GYR+Y DD+V F+ K+ ENGIDI RIFDALND R
 Sbjct: 62 ERLRALKKRIQNTPLQMLLRGQNLVGYRHYPPDIVEKFKVIKAHENGIDIFRIFDALNDVR 121
 55 Query: 130 NLQTAVSATKKFGGHAQVAISYTTSPVHTIDYFVELAKAYQAIGADSIKIDMAGVLTPE 189
 N++TA+ KK G Q AI YT SPVHTID +VELAK + +G DSICIKDMAG+LTP
 Sbjct: 122 NMETAIKTAKKVGAEVQGAICYTISPVHTIDQYVELAKKLEEMGCDSICIKDMAGLITPY 181

-2781-

Query: 190 IGYQLVKCIKENTTTIPLEVHTHATSGISEMTYLVKVAEAGADIIDTAISSFSGGTSQPATE 249
 GY+LVK +KE ++P++VH+H TSG++ MTYLVK EAGAD++D AIS F+ GTSQP TE
 Sbjct: 182 EGYELVKRLKEEISLPIDVHSHCTSG LAPMTYLVKVEAGADMVDCAISPFAMGTSQPPTTE 241

Query: 250 SMAIALTDLGFDTGLDMQEVAKVAEYFNTIRDHYREIGILNPKVKDTEPKTLIYQVPGGM 309
 S+ +AL +DTGLD++ + ++ +YF +R+ Y+ + +P + + + L+YQVPGGM
 Sbjct: 242 SIVVALKGT KYDTGLDLKLLNEIRDYFMKVREKYKM--LFSPISQIVDARVLVYQVPGGM 299

Query: 310 LSNLLSQLTEQGLTDKYEEVLAEVPKVRADLGYPPPLVTPLSQMVGTQALMNIISGERYKV 369
 LSNL+SQL EQG DK+EEVL E+P+VR DLGYPPPLVTP SQ+VGTQA++N+++ ERYK+
 Sbjct: 300 LSNLVSQLEQGALDKFEEVLQEIPRVKDLGYPPPLVTPTSQIVGTQAVLNVLTEERYKI 359

Query: 370 VPNEIKDYVRGLYGQSPAPLAEGIKEKIIGD-EAVITCRPADLIEPQMIYLRDEIAP--Y 426
 + NE+ +YV+G YG+ PAP+ + +++++ + E ITCRPADL+ P+ ++ E
 Sbjct: 360 ITNEVVNYVKGFGYKPPAPINPELLKRVLDEGEKPITCRPADLLPFEWEKVKKEAEEKGI 419

Query: 427 AHSEEDVLSYASFPQQARDFL 447
 EED+L+YA +PQ A FL
 Sbjct: 420 VKKEEDILTYALYPQIAVKFL 440

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2728

A DNA sequence (GASx1185R) was identified in *S.pyogenes* <SEQ ID 7955> which encodes the amino acid sequence <SEQ ID 7956>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2497(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF93960 GB:AE004165 citrate (pro-3S)-lyase ligase [Vibrio cholerae]
 Identities = 118/336 (35%), Positives = 183/336 (54%), Gaps = 5/336 (1%)

Query: 4 YTISKVFPSDKTTMASVKNNLHQQEGIRLDAHLDTCAIMNAQNDVIATGSYFGNSLRCLC 63

YT S+V ++T + +K L Q + +D +++ + N +IA G G+ L+ +

Sbjct: 10 YTFSRVSTKNRTKLLQIKEFLCQHQLTVDDDDVEHF-VVAYGTNQTIIACGGIAGHVLKSIA 68

Query: 64 VSSAYQGEGLLNRIVSHLIDEEYALGNHYLFVYTKTSSAAFFKDLGFTEIVHIDNHISFL 123

VS A QG G +++++ L + Y +G + LF++TK ++ F+ GF + ++ HI+ L

Sbjct: 69 VSPALQGTGFALKLMTLTFNAYEMGRFSLFLFTK PANIDLFRQCGFFLVDPKVEPHIAL 128

Query: 124 ENKKTGFQDYLMTLNKPEQTPGKVAIVINANPFTLGHQFLVEKAARENDWVHLFMVSED 183

EN Y L + + K+ +IV+NANPFTLGHQ+L+E+A + DWVHLF+V +

Sbjct: 129 ENSPNRLSVYCKQLQLLKMSGRKIGSIVMNANPFTLGHQYLIEQACEQCDWVHLFVVKA 188

Query: 184 RSLIPFSVRKRLIQEGLAHLNDNVIYHETGPYLISQATFPAYFQKEDNDVIKSQALLDTAI 243

++ R +I+ G HL N+ H Y+IS+ATFP+YF K+ V +S LD +I

Sbjct: 189 NKDFSADRMAMIKAGSKHLNLTIHSGSDYIISRATFP SYFIKDQQVNVQSHTALDL 248

Query: 244 FL-KIAQTLQITKRYVGEEPTSRVTAIYNEIM--AEQLQQAGILLDILPRKAINQQODP 299

F IA L IT R+VG EP VT YN+ M E+ A + ++ + Q P

Sbjct: 249 FRHSIAPALGITHRFVGSEPICTVTRHYNQAMRRWLEEAHDASAPIQVVEIERSQQASQP 308

Query: 300 ISASTARQALKDNDWDLAKLLPKTSLDYFCSLKAQ 335

-2782-

ISAS R LK + +A L+PKT+ Y C A+
 Sbjct: 309 ISASRVRYLLKQFGFAALADLVPKTTYSLCQHYAE 344

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2729

A DNA sequence (GASx1187) was identified in *S.pyogenes* <SEQ ID 7957> which encodes the amino acid sequence <SEQ ID 7958>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4790 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2730

A DNA sequence (GASx1188R) was identified in *S.pyogenes* <SEQ ID 7959> which encodes the amino acid sequence <SEQ ID 7960>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3956 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2731

A DNA sequence (GASx1190) was identified in *S.pyogenes* <SEQ ID 7961> which encodes the amino acid sequence <SEQ ID 7962>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1274 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2732

A DNA sequence (GASx1196R) was identified in *S.pyogenes* <SEQ ID 7963> which encodes the amino acid sequence <SEQ ID 7964>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2733

A DNA sequence (GASx1211) was identified in *S.pyogenes* <SEQ ID 7965> which encodes the amino acid sequence <SEQ ID 7966>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1850(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2734

A DNA sequence (GASx1219R) was identified in *S.pyogenes* <SEQ ID 7967> which encodes the amino acid sequence <SEQ ID 7968>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2284(Affirmative) < succ>

-2784-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

- 5 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2735

- 10 A DNA sequence (GASx1225) was identified in *S.pyogenes* <SEQ ID 7969> which encodes the amino acid sequence <SEQ ID 7970>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2062(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2736

- 25 A DNA sequence (GASx1229) was identified in *S.pyogenes* <SEQ ID 7971> which encodes the amino acid sequence <SEQ ID 7972>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

- 30

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2755(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 35

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

- 40 **Example 2737**

A DNA sequence (GASx1247R) was identified in *S.pyogenes* <SEQ ID 7973> which encodes the amino acid sequence <SEQ ID 7974>. Analysis of this protein sequence reveals the following:

Possible site: 31

- 45

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -6.32 Transmembrane 55 - 71 (53 - 81)

-2785-

```

INTEGRAL    Likelihood = -6.00    Transmembrane    74 - 90 ( 72 - 95)
INTEGRAL    Likelihood = -2.18    Transmembrane    95 - 111 ( 95 - 111)
INTEGRAL    Likelihood = -1.54    Transmembrane    124 - 140 ( 123 - 141)

```

```

5  ----- Final Results -----
      bacterial membrane --- Certainty=0.3527(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

10 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAB14326 GB:Z99116 yqjA [Bacillus subtilis]
Identities = 97/306 (31%), Positives = 154/306 (49%)

```

```

15 Query: 6   RTLKMTLATIVAILIAYQLHLDYAMSAGIIALLSVLDTRKSSLVARNRLLSFFFLAFGIA 65
      RT+K  L T +AI I+  LHL      SAGII +L +  T+K SL +  R +  LA  +
Sbjct: 7   RTIKTALGTALAIYISQLLHLQNFASAGIITILCIQITQKRSLQASWARFWACCLAIAFS 66

20 Query: 66  MMCFSLFGFTTVGFMCYLLIIIPLLYHFQIEAGLVPITVLVTHLIAKKSIALPILSNEFM 125
      + F L G+      LLI IP+   +I  G+V +V++ HL      I      + NE
Sbjct: 67  YLFFELIGYHPFVIGALLLIFIPITVLLKINEGIVTSSVILHLYMSGGITPTFIWNEVQ 126

25 Query: 126 LFFVGTSTVALLFNAYMGPDQQIRYYHQKVESDLKGILYRFESFLLEGKGQNEGLLIKNL 185
      L  VG  VALL N YM  D+++ Y +K+E +   I   E +LL G+   G  I
Sbjct: 127  LITVGIGVALLMNLYMPSLDRKLIAYRKKIEDNFAVIFAEIERYLLTGEDWSGKEIPET 186

30 Query: 186 DKILDEALKLVYRERHNQLFQQTNYQVHYFEMRRQNRLLGQMAINVNTLMRQSKESILL 245
      +++ EA  L YR+  N + +  N  HYF+MR +Q  ++ ++  V ++      + ++
Sbjct: 187  HQLITEAKNLAYRDVQNHLIRYENLHYHYFKMREKQFEIIERLLPKVTSISITVDQGKMI 246

35 Query: 246 SHLFHETACQLSEQNPALTLLIDDIEQLLETFRHGDLPTREEFERRAVLFQLQDLERFI 305
      +  H+      +  N A  +  +  +  F   LP TREEFE RA LF LL ++E+++
Sbjct: 247  AEFIHDLREAIHPGNTAYKFLKRLADMRKEFEEMPLPATREEFEARAALFHLGEMEQL 306

Query: 306 LLKVEF 311
      ++K  F
Sbjct: 307 VIKSYF 312

```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
40 antigens for vaccines or diagnostics.

Example 2738

A DNA sequence (GASx1261) was identified in *S.pyogenes* <SEQ ID 7975> which encodes the amino acid
sequence <SEQ ID 7976>. Analysis of this protein sequence reveals the following:

```

Possible site: 15
45 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.6082(Affirmative) < succ>
50      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

55 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
antigens for vaccines or diagnostics.

Example 2739

A DNA sequence (GASx1262R) was identified in *S.pyogenes* <SEQ ID 7977> which encodes the amino acid sequence <SEQ ID 7978>. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -7.06 Transmembrane 38 - 54 (37 - 55)

----- Final Results -----

bacterial membrane --- Certainty=0.3824(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2740

A DNA sequence (GASx1265R) was identified in *S.pyogenes* <SEQ ID 7979> which encodes the amino acid sequence <SEQ ID 7980>. Analysis of this protein sequence reveals the following:

Possible site: 25

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2741

A DNA sequence (GASx1270) was identified in *S.pyogenes* <SEQ ID 7981> which encodes the amino acid sequence <SEQ ID 7982>. Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4063(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2787-

Example 2742

A DNA sequence (GASx1290R) was identified in *S.pyogenes* <SEQ ID 7983> which encodes the amino acid sequence <SEQ ID 7984>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -12.37	Transmembrane	180 - 196 (172 - 207)
INTEGRAL	Likelihood = -10.19	Transmembrane	34 - 50 (30 - 53)
INTEGRAL	Likelihood = -4.09	Transmembrane	233 - 249 (232 - 250)

----- Final Results -----

bacterial membrane	---	Certainty=0.5946 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB88010 GB:L21856 Mala [Streptococcus pneumoniae]

Identities = 66/237 (27%), Positives = 105/237 (43%), Gaps = 28/237 (11%)

Query: 45 MIPVTLHYANMTTYPLERIVTKSLSPITDKTYQALTQGKIEKD---TFQGQSLIRRD--- 98

M+P+ + ++ TYPE + P+TDK Q L++ D T+ G +

Sbjct: 1 MVPIAIQNSSQETYPLETFIDNVYEPLTDKVVQDLSEHATIVDGTLTGTASQAPSVVI 60

Query: 99 GELVLAVLPTKVDLEQLASESTRQIIVTKKEWRFPDQKEL-RAHVRGQQQSLADLTIV 157

G + LP + L T +++++K + KEL R R Q T

Sbjct: 61 GPSQIKELPKDLQLHF----DTNELVISK-----ESKELTRISYRAIQ-----TEG 102

Query: 158 KAVKDFVNQQWY---DSNKASVLGFLLLTFVLMVCVGTLLIVIGLGAFFLTITKRSRLFMI 214

KD + Q + +N+ + FL+L + + IV L +TK+SRLF

Sbjct: 103 FKSKDSLTAQAFIRLVPTNRVYISLFLVLGASFLFGLNFFIVSLGACLLLYITKKSRLFSF 162

Query: 215 RNFSEGLGLMVNCLAWPSLLAIALSFFIQDPVLIMNCQVFGTLLMLTWVFYKTQFRD 271

R F E ++NCL P+L+ + L F Q+ ++ Q +L L +FYKT FRD

Sbjct: 163 RTFKECYHFILNCLGLPTLITLILGLFGQNMTTLITVQNILFVLYLVITIFYKTHFRD 219

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2743

A DNA sequence (GASx1294) was identified in *S.pyogenes* <SEQ ID 7985> which encodes the amino acid sequence <SEQ ID 7986>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.2104 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2788-

Example 2744

A DNA sequence (GASx1303R) was identified in *S.pyogenes* <SEQ ID 7987> which encodes the amino acid sequence <SEQ ID 7988>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -8.07 Transmembrane 13 - 29 (8 - 38)

----- Final Results -----

bacterial membrane --- Certainty=0.4227(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2745

A DNA sequence (GASx1307R) was identified in *S.pyogenes* <SEQ ID 7989> which encodes the amino acid sequence <SEQ ID 7990>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2746

A DNA sequence (GASx1312R) was identified in *S.pyogenes* <SEQ ID 7991> which encodes the amino acid sequence <SEQ ID 7992>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1996(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2789-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2747

A DNA sequence (GASx1316R) was identified in *S.pyogenes* <SEQ ID 7993> which encodes the amino acid sequence <SEQ ID 7994>. Analysis of this protein sequence reveals the following:

Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3504(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 271-273

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC66321 GB:AE000792 outer surface protein, putative [Borrelia burgdorferi]
Identities = 127/365 (34%), Positives = 195/365 (52%), Gaps = 14/365 (3%)

Query: 1 MVDLGFSLYPERYDVTKSKAYIDLCHSYGAKRLFMSLLQLAPADHQMFCYAEIAYANQ 60
M ++G S+YP K Y++ +G ++F SLL + + F + EL++ AN+
Sbjct: 1 MKEIGISITYPNVSPKNKIIKYLESAHFGFTQVFTSLLYI---NGNEFDIFKELLSIANK 57

Query: 61 LGIRVIADVSPSFISQAGWSDQLIERA-----HAFGLAGLRLEALPLAEIVTLTRNPF 114
G++ I DVSP + G + G +RLD E +T N
Sbjct: 58 NGMKPIIDVSPEIFKELGIDLNLNRCPKLDYFKKLGAWAIRLDNTFTGIEESLMTFNDS 117

Query: 115 GLKIELNMSTDKQLLSLLATDAERSNIIGCHNFYPHEFTGLSWQHFKMSRFYHEHDIE 174
LKI+LN+S + + +++ N++GCHNFYPH++TGLS FK+ ++ + + I
Sbjct: 118 DLKIQNLNISNINKHIDTIMYFKPNIKNLLGCHNFYPHKYTGLSRNFFKETTKIFKHYSIP 177

Query: 175 TAAFTTAQSASE-GPWLLAEGLEPTVEDHRHLPGLQVELMKAIGTIDNILISNQFISEEE 233
TAAFI++ +A E EG+PT+E HR I Q + + G ID +LISN F SE E
Sbjct: 178 TAAFISSNNAEECARGKEKEGVPTLESHRSKDIETQAKDLFKEG-IDTVLISNCFPSETE 236

Query: 234 LAACTQALARPVTTIKVRPIIDLTEVEEQII-GYPHCYRGDVSDYVIRSTMPRLVYAQES 292
L ++ + R + +K D VE++II H RGD++ Y IRSTMPR+ Y +
Sbjct: 237 LKKVSK-VNRNILELKADLNPDANSVEKEIILENLHFNRGDINSYRIRSTMPRVYNNKK 295

Query: 293 IAPRQDSKEVKGSIIDNDRYHRYKGELQIALKNFTVSSKANVVAEVREDYLSLLDDL 352
P E+K+G I+ID+ Y Y GELQIALK+ + NVV ++ D + LL+ +
Sbjct: 296 F-PVHSPNEIKKGDILIDSSEYLGTYGELQIALKDTPNNGLVNVVGKIINDEIYLLKIE 354

Query: 353 PWQEF 357

PW++F

Sbjct: 355 PWEKF 359

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2748

A DNA sequence (GASx1319) was identified in *S.pyogenes* <SEQ ID 7995> which encodes the amino acid sequence <SEQ ID 7996>. Analysis of this protein sequence reveals the following:

Possible site: 34

-2790-

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.50	Transmembrane	127 - 143 (125 - 151)
INTEGRAL	Likelihood = -7.43	Transmembrane	17 - 33 (15 - 36)
INTEGRAL	Likelihood = -5.68	Transmembrane	39 - 55 (36 - 57)
INTEGRAL	Likelihood = -1.86	Transmembrane	60 - 76 (59 - 77)
INTEGRAL	Likelihood = -0.59	Transmembrane	85 - 101 (85 - 101)

----- Final Results -----

bacterial membrane --- Certainty=0.4800(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

- 15 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2749

A DNA sequence (GASx1320) was identified in *S.pyogenes* <SEQ ID 7997> which encodes the amino acid sequence <SEQ ID 7998>. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -1.81	Transmembrane	35 - 51 (35 - 51)
----------	--------------------	---------------	--------------------

----- Final Results -----

bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 30 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2750

- 35 A DNA sequence (GASx1321) was identified in *S.pyogenes* <SEQ ID 7999> which encodes the amino acid sequence <SEQ ID 8000>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2791-

Example 2751

A DNA sequence (GASx1329) was identified in *S.pyogenes* <SEQ ID 8001> which encodes the amino acid sequence <SEQ ID 8002>. Analysis of this protein sequence reveals the following:

Possible site: 44

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.28 Transmembrane 64 - 80 (64 - 80)

----- Final Results -----

bacterial membrane --- Certainty=0.1510 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2752

A DNA sequence (GASx1332R) was identified in *S.pyogenes* <SEQ ID 8003> which encodes the amino acid sequence <SEQ ID 8004>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2753

A DNA sequence (GASx1333) was identified in *S.pyogenes* <SEQ ID 8005> which encodes the amino acid sequence <SEQ ID 8006>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2792-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2754

A DNA sequence (GASx1335R) was identified in *S.pyogenes* <SEQ ID 8007> which encodes the amino acid sequence <SEQ ID 8008>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF96047 GB:AE004354 uridine phosphorylase [Vibrio cholerae]

Identities = 46/167 (27%), Positives = 72/167 (42%), Gaps = 12/167 (7%)

Query: 8 GVKEMISTGTCGVLVP-IAENRFLVPVKALRDEGTSYHYVAPSRVIDIDPKMLRLIEKTL 66
G K ++ G+ G + I ++ A+RDEG S Y+ + +++ ++ L
Sbjct: 79 GAKAIVRVGSAGAMQSEIGLGLILVEGAVRDEGGSKAYIGAAYPAYSSFELVVEMQRFL 138

Query: 67 LAQGLAYQEVITWSTDGFYR-ETKEKVAHRQEEGCSVVEMECSALAAVAQLRG-----IL 120
Q + I S D F Y E E + +G +ME SAL V +LRG +L
Sbjct: 139 AEQSVPIHRGIVRSHDSFYTDEEAELCRYWHRKGILAADMETSALLTVGRLRGLQVASVL 198

Query: 121 WGQLLFTADTLADVEVY---DQRNWGADSFSAHLHLCLEVLNTLEKD 164
+L+ D A V Y DQR + + A L LN L+ D

Sbjct: 199 NNVVLYEQDVQAGVNQYVNADQRMQGE--TLAARAALHALNALKFD 243

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2755

A DNA sequence (GASx1353) was identified in *S.pyogenes* <SEQ ID 8009> which encodes the amino acid sequence <SEQ ID 8010>. Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -5.79 Transmembrane 241 - 257 (234 - 260)
INTEGRAL Likelihood = -5.15 Transmembrane 44 - 60 (43 - 65)
INTEGRAL Likelihood = -4.78 Transmembrane 74 - 90 (72 - 92)

----- Final Results -----

bacterial membrane --- Certainty=0.3314 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2793-

Example 2756

A DNA sequence (GASx1354R) was identified in *S.pyogenes* <SEQ ID 8011> which encodes the amino acid sequence <SEQ ID 8012>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -3.45 Transmembrane 68 - 84 (65 - 86)

----- Final Results -----

bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB83831 GB:AL162753 putative integral membrane protein
[Neisseria meningitidis]

Identities = 31/72 (43%), Positives = 46/72 (63%), Gaps = 6/72 (8%)

Query: 17 FVIYAFDKRKAIAKKRRISERKLLVITVLFGGF-GALLAAKKYHHKTRKWFVI----TC 71
F +Y DKR+A++ KRRI E +LL + LFGG+ GA L ++ + HKT K FV+ T
Sbjct: 38 FALYGIDKRRRAVRGKRRRIPEHRL-LPALFGGWAGAYLGSRIFRHKTAKKRFVVLFRITV 96

Query: 72 YTSILLTLLVTY 83
++L TL++ Y

Sbjct: 97 SGNVLATLILLY 108

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2757

A DNA sequence (GASx1363R) was identified in *S.pyogenes* <SEQ ID 8013> which encodes the amino acid sequence <SEQ ID 8014>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2758

A DNA sequence (GASx1367) was identified in *S.pyogenes* <SEQ ID 8015> which encodes the amino acid sequence <SEQ ID 8016>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have an uncleavable N-term signal seq

-2794-

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:CAA63508 GB:X92946 hypothetical protein [Lactococcus lactis]
 Identities = 64/96 (66%), Positives = 77/96 (79%)

 Query: 1 MPRKTFDKAFKLSAVKLIILEEQPVKMVSSTLEIHPNSLYQWIQEYKYGESAFFPGHGSA 60
 M R+ FDK FK SAVKLILEE VK VS LE+H NSLY+W+QE E+YGESAFFPG+G+A
 Sbjct: 1 MARRKFDKQFKNSAVKLIILEEGYSVKEVSQELEVHANSLYRWVQEVEEYGESAFFPGNGTA 60

 Query: 61 LRHAQFKTKKLEKEHKLLQEELALLKKFQVFLKPNR 96
 L +AQ K K LEKE++ LQEEL LLKKF+VFLK ++
 Sbjct: 61 LANAQHKIKLLEKENRYLQEELALLKKFRVFLKRSK 96

 Query: 61 LRHAQFKTKKLEKEHKLLQEELALLKKFQVFLKPNR 96
 L +AQ K K LEKE++ LQEEL LLKKF+VFLK ++
 Sbjct: 61 LANAQHKIKLLEKENRYLQEELALLKKFRVFLKRSK 96

20 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2759

A DNA sequence (GASx1374R) was identified in *S.pyogenes* <SEQ ID 8017> which encodes the amino acid sequence <SEQ ID 8018>. Analysis of this protein sequence reveals the following:

25 Possible site: 39

 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2585(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

35 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2760

40 A DNA sequence (GASx1382R) was identified in *S.pyogenes* <SEQ ID 8019> which encodes the amino acid sequence <SEQ ID 8020>. Analysis of this protein sequence reveals the following:

 Possible site: 14

 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -2.39 Transmembrane 3 - 19 (3 - 19)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

-2795-

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2761

- 5 A DNA sequence (GASx1391R) was identified in *S.pyogenes* <SEQ ID 8021> which encodes the amino acid sequence <SEQ ID 8022>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> May be a lipoprotein

10

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

15

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 2762

A DNA sequence (GASx1404) was identified in *S.pyogenes* <SEQ ID 8023> which encodes the amino acid sequence <SEQ ID 8024>. Analysis of this protein sequence reveals the following:

Possible site: 32

25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3046 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

30

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

- 35 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2763

A DNA sequence (GASx1412R) was identified in *S.pyogenes* <SEQ ID 8025> which encodes the amino acid sequence <SEQ ID 8026>. Analysis of this protein sequence reveals the following:

Possible site: 20

40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1590 (Affirmative) < succ>

45

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-2796-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2764

A DNA sequence (GASx1414R) was identified in *S.pyogenes* <SEQ ID 8027> which encodes the amino acid sequence <SEQ ID 8028>. Analysis of this protein sequence reveals the following:

Possible site: 24

10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2816(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

20 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2765

A DNA sequence (GASx1416) was identified in *S.pyogenes* <SEQ ID 8029> which encodes the amino acid sequence <SEQ ID 8030>. Analysis of this protein sequence reveals the following:

Possible site: 34

25 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.1744(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

35 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2766

A DNA sequence (GASx1417) was identified in *S.pyogenes* <SEQ ID 8031> which encodes the amino acid sequence <SEQ ID 8032>. Analysis of this protein sequence reveals the following:

40 Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.3771(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-2797-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2767

A DNA sequence (GASx1419R) was identified in *S.pyogenes* <SEQ ID 8033> which encodes the amino acid sequence <SEQ ID 8034>. Analysis of this protein sequence reveals the following:

10 Possible site: 13

>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -10.93 Transmembrane 4 - 20 (1 - 25)

15 ----- Final Results -----

bacterial membrane --- Certainty=0.5373(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 20 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2768

- 25 A DNA sequence (GASx1423) was identified in *S.pyogenes* <SEQ ID 8035> which encodes the amino acid sequence <SEQ ID 8036>. Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -8.97 Transmembrane 30 - 46 (25 - 49)
 INTEGRAL Likelihood = -7.80 Transmembrane 52 - 68 (50 - 72)
 INTEGRAL Likelihood = -6.95 Transmembrane 129 - 145 (125 - 146)

35 ----- Final Results -----

bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

- 40 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2769

- 45 A DNA sequence (GASx1426R) was identified in *S.pyogenes* <SEQ ID 8037> which encodes the amino acid sequence <SEQ ID 8038>. Analysis of this protein sequence reveals the following:

Possible site: 25

-2798-

>>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -3.45 Transmembrane 36 - 52 (36 - 55)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC39287 GB:AF115103 orf87 gp [Streptococcus thermophilus
 bacteriophage Sfi21]
 Identities = 43/73 (58%), Positives = 61/73 (82%)
 15 Query: 1 MINLKLRLQNKVTLMAILGAIFLLAQQLGKLPNSNIADIANAVTLVLGVTDPPTKG 60
 MIN KLRLQNK TL+A++ A+FL+ QQ G+ +P+NI + NT V +LV+LG++TDPTTKG
 Sbjct: 8 MINFKLRLQNKATLVALISAVFLMLQQFGLHVPNNIQEGINTLVGILVILGIITDPTTKG 67
 20 Query: 61 LSDSEQALTYHEP 73
 ++DSE+AL+Y +P
 Sbjct: 68 IADSERALSIIQP 80

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 25 antigens for vaccines or diagnostics.

Example 2770

A DNA sequence (GASx1427R) was identified in *S.pyogenes* <SEQ ID 8039> which encodes the amino acid sequence <SEQ ID 8040>. Analysis of this protein sequence reveals the following:

Possible site: 27
 30 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -3.03 Transmembrane 2 - 18 (1 - 23)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

40 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2771

45 A DNA sequence (GASx1428R) was identified in *S.pyogenes* <SEQ ID 8041> which encodes the amino acid sequence <SEQ ID 8042>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence
 50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1017(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2799-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
5 antigens for vaccines or diagnostics.

Example 2772

A DNA sequence (GASx1429R) was identified in *S.pyogenes* <SEQ ID 8043> which encodes the amino acid sequence <SEQ ID 8044>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3097(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
20 antigens for vaccines or diagnostics.

Example 2773

A DNA sequence (GASx1431R) was identified in *S.pyogenes* <SEQ ID 8045> which encodes the amino acid sequence <SEQ ID 8046>. Analysis of this protein sequence reveals the following:

Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2584(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA98101 GB:M19348 hyaluronidase [Streptococcus pyogenes phage
H4489A]

Identities = 337/371 (90%), Positives = 351/371 (93%), Gaps = 1/371 (0%)

Query: 1 MAENIPLRVQFKRMKAAEWASSDVVLLEGEIGFETDTGFAKFGDGQNTFSKLYLTGPKG 60

M ENIPLRVQFKRM A EWA SDV+LLEGEIGFETDTGFAKFGDGQNTFSKLYLTGPKG

Sbjct: 1 MTENIPLRVQFKRMSADEWARSVDVILLEGEIGFETDTGFAKFGDGQNTFSKLYLTGPKG 60

Query: 61 PKGDTGLQGKTGGTGSRGPAKPGTTDQDLQNKPDLGAFQAQKEETNSKITKLESSKADK 120

PKGDTGLQGKTGGTG RGPAGKPGTTDQDLQNKPDLGAFQAQKEETNSKITKLESSKADK

Sbjct: 61 PKGDTGLQGKTGGTGPRGPAGKPGTTDQDLQNKPDLGAFQAQKEETNSKITKLESSKADK 120

Query: 121 NAVYLKAESNAKLDEKLNKGGVMTGQLQFKPN-SGIKPSSSVGGAINIDMSKSEGAAMV 179

+AVY KAES +LD+KL+L GG++TGQLQFKPN SGIKPSSSVGGAINIDMSKSEGAAMV

Sbjct: 121 SAVYKAESKIELDKKLSLTGGIVTGQLQFKPNKSGIKPSSSVGGAINIDMSKSEGAAMV 180

-2800-

Query: 180 MYTNKDTTGDGPLMILRSNKDTFDQSVQFVDYKGTNAVNIVMRQPTTPNFSSALNITSAN 239
 MYTNKDTTGDGPLMILRS+KDTFDQS QFVDY G TNAVNIVMRQP+ PNFSSALNITSAN
 Sbjct: 181 MYTNKDTTGDGPLMILRSKDTFDQSAQFVDYSGKTNAVNIVMRQPSAPNFSSALNITSAN 240

5 Query: 240 EGGSAMQIRGVEKALGTLKITHENPSVDKEYDENAAALSIDIVKKQGGKGTAAGGIYIN 299
 EGGSAMQIRGVEKALGTLKITHENP+V+ +YDENAAALSIDIVKKQGGKGTAAGGIYIN
 Sbjct: 241 EGGSAMQIRGVEKALGTLKITHENPNVEAKYDENAAALSIDIVKKQGGKGTAAGGIYIN 300

10 Query: 300 STSGTAGKMLRIRNKNKDKFYVGPDPGDFWSCASSIVDGNLTVKDPTSGKHAATKDYVDEK 359
 STSGTAGKMLRIRNKN+DKFYVGPDPG F S A+S V GNLTVDPTSGKHAATKDYVDEK
 Sbjct: 301 STSGTAGKMLRIRNKNEDKFYVGPDPGGFHSGANSTVAGNLTVKDPTSGKHAATKDYVDEK 360

Query: 360 IAEKKKLILKK 370
 IAEKKKLILKK
 15 Sbjct: 361 IAEKKKLILKK 371

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2774

20 A DNA sequence (GASx1438R) was identified in *S.pyogenes* <SEQ ID 8047> which encodes the amino acid sequence <SEQ ID 8048>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

25

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30

A related DNA sequence <SEQ ID 10439> was identified in GBS which encodes amino acid sequence <SEQ ID 10440>.

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

35 >GP:AAB18711 GB:U38906 ORF36 [Bacteriophage rlt]
 Identities = 70/111 (63%), Positives = 88/111 (79%)

Query: 1 LIEVIKKYLDEHLDVPSFFEHQDEPARFIILEKTSQAKQNHLLSSTFAFQSYAESLYE 60
 +IE+IIK +LD HL V SF E + + P +I+ EKT +K NHLLSSTFAFQSYA S+YE
 40 Sbjct: 1 MIEIIKNFLDTHLSVSSFLEKKGEMPLSYILFEKTGSSKSNHLLSSTFAFQSYAPSMYE 60

Query: 61 AALLNDKVKQVIEQLDVLQPQVSGVHLNADYNFTDTATKRYRYQAVFDINHY 111
 AA LN+++K+V+E+L L ++S V LN+DYNFTDT TK YRYQAVFDINHY
 45 Sbjct: 61 AAKLNEQLKEVVERLIELNEISNVSLNSDYNFTDTETKEYRYQAVFDINHY 111

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2775

50 A DNA sequence (GASx1442R) was identified in *S.pyogenes* <SEQ ID 8049> which encodes the amino acid sequence <SEQ ID 8050>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

-2801-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1241(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

10 Example 2776

A DNA sequence (GASx1444R) was identified in *S.pyogenes* <SEQ ID 8051> which encodes the amino acid sequence <SEQ ID 8052>. Analysis of this protein sequence reveals the following:

Possible site: 42

15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4547(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25

Example 2777

A DNA sequence (GASx1447R) was identified in *S.pyogenes* <SEQ ID 8053> which encodes the amino acid sequence <SEQ ID 8054>. Analysis of this protein sequence reveals the following:

Possible site: 25

30

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40

Example 2778

A DNA sequence (GASx1448R) was identified in *S.pyogenes* <SEQ ID 8055> which encodes the amino acid sequence <SEQ ID 8056>. Analysis of this protein sequence reveals the following:

Possible site: 20

45

-2802-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3221(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

10 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2779

A DNA sequence (GASx1449R) was identified in *S.pyogenes* <SEQ ID 8057> which encodes the amino acid sequence <SEQ ID 8058>. Analysis of this protein sequence reveals the following:

15 Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.6356(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

25 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2780

30 A DNA sequence (GASx1453R) was identified in *S.pyogenes* <SEQ ID 8059> which encodes the amino acid sequence <SEQ ID 8060>. Analysis of this protein sequence reveals the following:

 Possible site: 13

>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.2869(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2781

45 A DNA sequence (GASx1455R) was identified in *S.pyogenes* <SEQ ID 8061> which encodes the amino acid sequence <SEQ ID 8062>. Analysis of this protein sequence reveals the following:

-2803-

Possible site: 40

>>> Seems to have no N-terminal signal sequence

5 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1787(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

10 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF43512 GB:AF145054 ORF19 [Streptococcus thermophilus
 bacteriophage 7201]
 Identities = 47/126 (37%), Positives = 86/126 (67%), Gaps = 2/126 (1%)

15 Query: 8 LKDLRNLDLYIASLIRRRDKIEASLL--SSPKWSSDKVNGGIKQDDVYVELIATAKDI 65
 ++ ++ LD YI S I + ++E+ L +S +D V GG ++ +DD+YVELI +++
 Sbjct: 7 IQQIKALDRYIESQIEQIKRLESQALKVTSGSMHTDMVQGGKRGKDDIYVELITAREEV 66

20 Query: 66 EKKTAEAIRKQRELQNLIDSLENTDSQTILSMVYIDKMTRWQVIDELNCSESTYFRLLRV 125
 E+ TAEAI+++ E + I ++E+ D++++L MVYID+++ WQ+ D++ S++TY+ LR
 Sbjct: 67 ERFTAEAIAKQKLEFRQIANIEDIDARSLLQMVYIDQLSIWQICDKMGISKATYYVKLRQ 126

25 Query: 126 ATKEIN 131
 A K L+
 Sbjct: 127 AEKYLD 132

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 2782

A DNA sequence (GASx1456R) was identified in *S.pyogenes* <SEQ ID 8063> which encodes the amino acid sequence <SEQ ID 8064>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2883(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB18697 GB:U38906 ORF22 [Bacteriophage rlt]
 Identities = 78/207 (37%), Positives = 123/207 (58%), Gaps = 2/207 (0%)

45 Query: 6 EIHRIIGIDEVYKAPKRLTDILFDKDSREDIFRQFLKYETDVSYDWFQMIFYEEEQADRKN 65
 + + +L +DE R+ +++FDK RE+ + + L D+ D+F YF A
 Sbjct: 7 QFYDMLNVDEHMFTNRIQELVFDKKGREEFYSKILNIHDMGVDFFRDYFMAHSAVSA- 65

50 Query: 66 KKQDFTPKSVSTLLSKIISGNQYYEVA-VGTGGILIQAWQEQRLNDSPTFYRPSKYWYHV 124
 K Q +TP + L + ++ G+ ++ GTG ++IQ WQ+ R+N F Y PS YWY
 Sbjct: 66 KGQHYTPDELGKLTALLVGGSGGADLTGAGTGTLIIQKWQDDRMNTDFFNYPNYSNYWYQA 125

55 Query: 125 EELSDKAVPFLLFNMSIRGINGVVVHGDLSLRQVKNIYFLQNTKDDMLSFSNDINVPRTQ 184
 ELSD+A+ FL+ +IRG+NGVV+HGD+L VK +YF+QN+ ++ + FS+INV+P ++
 Sbjct: 126 LELSDEAISFLIHAFIRGMNGVVIHGDALMAVKQVYFIQNSANNPIGFSEINVIPHSK 185

-2804-

Query: 185 DIREFNVKEWIGDIEHIENPLIEWI 211
 D + EW IEIE+ +WI
 Sbjct: 186 DAMEFLGIHEWTEQAIEHIESKFPDWI 212

- 5 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2783

A DNA sequence (GASx1459R) was identified in *S.pyogenes* <SEQ ID 8065> which encodes the amino acid sequence <SEQ ID 8066>. Analysis of this protein sequence reveals the following:

10 Possible site: 16

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.44 Transmembrane 82 - 98 (81 - 98)

15 ----- Final Results -----
 bacterial membrane --- Certainty=0.1977(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 20 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2784

- 25 A DNA sequence (GASx1460R) was identified in *S.pyogenes* <SEQ ID 8067> which encodes the amino acid sequence <SEQ ID 8068>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3368(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 2785

A DNA sequence (GASx1461R) was identified in *S.pyogenes* <SEQ ID 8069> which encodes the amino acid sequence <SEQ ID 8070>. Analysis of this protein sequence reveals the following:

Possible site: 61

45 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2834(Affirmative) < succ>

-2805-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

- 5 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2786

- 10 A DNA sequence (GASx1462R) was identified in *S.pyogenes* <SEQ ID 8071> which encodes the amino acid sequence <SEQ ID 8072>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3531(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2787

- 25 A DNA sequence (GASx1463R) was identified in *S.pyogenes* <SEQ ID 8073> which encodes the amino acid sequence <SEQ ID 8074>. Analysis of this protein sequence reveals the following:

Possible site: 58

>>> Seems to have no N-terminal signal sequence

- 30 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2483(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 35 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14569 GB:Z99117 similar to phage-related protein [Bacillus subtilis]
 Identities = 98/252 (38%), Positives = 152/252 (59%), Gaps = 29/252 (11%)

- 40 Query: 16 SPAVKNRIEQVVGARAEQFTTSLLSIISNNLLAKATSESIMGAAMKAAVLNLPIEPSLG 75
 SP+V R E+V+G RA QFT S+LS+ ++ +L K S++ +AM AA L+LPI+ +LG
 Sbjct: 33 SPSVIKRFEEVLGKRATQFTASILSLYNSQMLQKTDPMSSVSSAMVAATLDLPIDKNLG 92

- 45 Query: 76 FAYVVPYNNRNYKDGNRWITVNEAQFQIGYRGLIQLAQRSGQVRNIEHGIIEEEFLGYDK 135
 +A++VPY +AQFQ+GY+G IQLA R+GQ ++I I+E E ++
 Sbjct: 93 YAWIVPYG-----GKAQFQLGYKGYIQLALRTGQYKSINCIPIHEGELQKWNP 140

- 50 Query: 136 IRGQLKLTGDYVDGSGVVGKGYFASLELISGFYKMIFWPKEKVEYEHAKKYSKTFDKKTGDFK 195
 + +++++ + +S V GY A ELI+GF K ++W K +V +H KK+SK+ DF
 Sbjct: 141 LTEEIEIDFEKRESDAVIGYAAYFELINGFRKTVYWTKAQVEKHKKKFSKS-----DF- 193

-2806-

Query: 196 PGTWPATEFDPMAIKTLLKELLISKYAPLSVEMQDA-LEADNADSTIVIPKDVTPQETNSI 254
 W ++D MA+KT+LK +LSK+ LSVEMQ A +E D I D+T + +S
 Sbjct: 194 ---GWKNDWDAMALKTVLKAVALSKWGLSVEMQKAVIEEDETRERI----DITNEADSS- 245

Query: 255 DDLIGTQNEKKD 266
 ++I ++ KD
 Sbjct: 246 -EIIDSEPSNKK 256

- 10 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2788

A DNA sequence (GASx1464R) was identified in *S.pyogenes* <SEQ ID 8075> which encodes the amino acid sequence <SEQ ID 8076>. Analysis of this protein sequence reveals the following:

15 Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.4258 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

- 25 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2789

A DNA sequence (GASx1465R) was identified in *S.pyogenes* <SEQ ID 8077> which encodes the amino acid sequence <SEQ ID 8078>. Analysis of this protein sequence reveals the following:

30 Possible site: 51

>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2045 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 40 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2790

- 45 A DNA sequence (GASx1469R) was identified in *S.pyogenes* <SEQ ID 8079> which encodes the amino acid sequence <SEQ ID 8080>. Analysis of this protein sequence reveals the following:

Possible site: 19

-2807-

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

5 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

10 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2791

A DNA sequence (GASx1470R) was identified in *S.pyogenes* <SEQ ID 8081> which encodes the amino acid sequence <SEQ ID 8082>. Analysis of this protein sequence reveals the following:

15 Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.3577 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

25 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC98430 GB:L29324 excisionase [Streptococcus pneumoniae]
 Identities = 23/56 (41%), Positives = 41/56 (73%)

30 Query: 23 KHLIQQWEGTLVATAKQWATEMRDHPDFKQFVLNPTHRIVFIDYKGFKLFVQWKSR 78
 K ++++W+GL T +W EMR++ F +V+NPTH++VFI+ +GF+ F++WK +
 Sbjct: 19 KGILKRWDLNKKYTLNRWIKEMRENRTFSMYVINPTHKLIVFINLEGFESFLRWKQK 74

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 Example 2792

A DNA sequence (GASx1473) was identified in *S.pyogenes* <SEQ ID 8083> which encodes the amino acid sequence <SEQ ID 8084>. Analysis of this protein sequence reveals the following:

Possible site: 27

40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2725 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

50 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2793

A DNA sequence (GASx1476) was identified in *S.pyogenes* <SEQ ID 8085> which encodes the amino acid sequence <SEQ ID 8086>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 23

   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.1422(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

15 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2794

A DNA sequence (GASx1480R) was identified in *S.pyogenes* <SEQ ID 8087> which encodes the amino acid sequence <SEQ ID 8088>. Analysis of this protein sequence reveals the following:

```

20   Possible site: 25

   >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -4.04    Transmembrane    291 - 307 ( 290 - 309)

25   ----- Final Results -----
      bacterial membrane --- Certainty=0.2614(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

30 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2795

35 A DNA sequence (GASx1489R) was identified in *S.pyogenes* <SEQ ID 8089> which encodes the amino acid sequence <SEQ ID 8090>. Analysis of this protein sequence reveals the following:

```

   Possible site: 23

   >>> Seems to have no N-terminal signal sequence

40   ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2278(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
45

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2809-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2796

A DNA sequence (GASx1490R) was identified in *S.pyogenes* <SEQ ID 8091> which encodes the amino acid sequence <SEQ ID 8092>:

```
SFITSVLAFRKLLKCEGIDLYLMYGDLMTCFEQLLTQLKDWTVDVYFNYDESGYGRLRDQKAAQFFKKNGLAVHTYQDHYLHGSQEIIINQSG
QPYKVFTPYRIWQNYPKETPIKVELSQGRWLNLETPDDVLRITVESFKDEKYQDVATFDEASKQLNRFIQDQLAAYHANRDFPAQLGTSRL
SPFLRIGAIGIRTVYHAVRQAPNSLGQATFLKELAWRDFYNMYYVAYPDQKTQPIQKAFSQIEWVNNPDWFQWLKKGKTYPIVDAAMLQL
QKTGWMHNRLRMIVASFLT KDLLCDWRLGEQYFQQQLIDYDAASNIGGWQWAASTGTDVAVPYFRIFNPVTQGKRFPKGEFIKAYLPQLEH
VPEKYLHEPWKMPKNLQESVSCIIGTDYPQPIVDHAKQREQAIAKYEWAKEKAKIE
```

Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA22361 GB:M94110 DNA photolyase [Bacillus firmus]
Identities = 175/338 (51%), Positives = 228/338 (66%), Gaps = 6/338 (1%)

Query: 145 EIINQSGQPYKVFTPYRIWQNYPKETP--IKVELSQGRWLNLETPDDVLRITVES--FKD 200
+++ + G PYKVFTPY+ W K TP IK ++ G PD T+ + K
Sbjct: 2 QVLKKGDTPIYKVFTPYKAWAKERKRTPAVIKRDVLLGSHVHKGTAEDREAEITLNNLIKK 61

Query: 201 EKYQDVATFDE-ASKQLNRFIQDQLAAYHANRDFPAQLGTSRLSPFLRIGAIGIRTVY-H 258
Y A +E A K+L F + +L+ Y ANRDFP+ GTSRLSP+++ GA+ R++Y H
Sbjct: 62 CSYDWSAIGEEHAIKRLQMFTKRLSGYKANRDFPSITGTSRLSPYIKTGAVSSRSIYYH 121

Query: 259 AVRQAPNSLGQATFLKELAWRDFYNMYYVAYPDQKTQPIQKAFSQIEWVNNPDWFQWLKE 318
+ +S TFLKELAWRDFY MV+ PD K + I + + ++ W ++ D WK
Sbjct: 122 ILNAEADSYSAETFLKELAWRDFYRMVHFYEPDCKDREIMEGYRELNWSHDQDDLTWKR 181

Query: 319 GKTGYPIVDAAMLQLQKTGWMHNRLRMIVASFLT KDLLCDWRLGEQYFQQQLIDYDAASN 378
G+TG+PIVDA M QL GWMHNRLRMI ASFLT KDLL DWRLGE+YF++ LIDYD +SN
Sbjct: 182 GETGFFIVDAGMRQLLNEGWMHNRLRMITASFLT KDLLIDWRLGERYFERMLIDYDPSSN 241

Query: 379 IGGWQWAASTGTDVAVPYFRIFNPVTQGKRFPKGEFIKAYLPQLEHVPEKYLHEPWKMPK 438
IGGWQWAAAS GTDAVPYFRIFNPVTQ KRFD G +I+ Y+P+L HVP+ Y+HEPWKM +
Sbjct: 242 IGGWQWAAASVGTDAVPYFRIFNPVTQSKRFDENGTYIRTYIPELNHVPDHYIHEPWKMSE 301

Query: 439 NLQESVSCIIGTDYPQPIVDHAKQREQAIAKYEWAKEK 476
Q C + DYP PIVDH+KQR++A++ ++ E+
Sbjct: 302 EEQVKYKCRLEDYPLPIVDHSHKQRKALSFYKGDDEE 339
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2797

A DNA sequence (GASx1493R) was identified in *S.pyogenes* <SEQ ID 8093> which encodes the amino acid sequence <SEQ ID 8094>. Analysis of this protein sequence reveals the following:

-2810-

Possible site: 39

>>> Seems to have no N-terminal signal sequence

5 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2748(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

10 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2798

15 A DNA sequence (GASx1501R) was identified in *S.pyogenes* <SEQ ID 8095> which encodes the amino acid sequence <SEQ ID 8096>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have a cleavable N-term signal seq.

20 INTEGRAL Likelihood = -7.27 Transmembrane 64 - 80 (53 - 83)

----- Final Results -----

25 bacterial membrane --- Certainty=0.3909(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

30 >GP:AAC95443 GB:AF068901 YlmG [Streptococcus pneumoniae]
 Identities = 35/81 (43%), Positives = 58/81 (71%)

Query: 1 MILILSILLRLIKVYTYLLIAYALMSWFPAYDSKIGRLISGIVEPILKPFRAFNLQFAG 60
 MI ++ ++ + +Y+ +L+A+A+MSWFPAY+S +GR I +V+P+L P + LQ AG
 Sbjct: 1 MIFLIRMIYNAVDIYSLILVAFVMSWFPAYESSLGRWIVALVKPVLAPLQRLPLQIAG 60

35 Query: 61 LDFTIFVVIISLNFLAQVLVR 81
 LD +++V I+ + FL + LVR
 Sbjct: 61 LDLSVWVAIVLVRFLGENLVR 81

40 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2799

A DNA sequence (GASx1502) was identified in *S.pyogenes* <SEQ ID 8097> which encodes the amino acid sequence <SEQ ID 8098>. Analysis of this protein sequence reveals the following:

45 Possible site: 25

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -2.39 Transmembrane 17 - 33 (17 - 33)

50 ----- Final Results -----

bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2811-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
5 antigens for vaccines or diagnostics.

Example 2800

A DNA sequence (GASx1507) was identified in *S.pyogenes* <SEQ ID 8099> which encodes the amino acid
sequence <SEQ ID 8100>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0865(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
20 antigens for vaccines or diagnostics.

Example 2801

A DNA sequence (GASx1511R) was identified in *S.pyogenes* <SEQ ID 8101> which encodes the amino
acid sequence <SEQ ID 8102>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -11.83 Transmembrane 31 - 47 (22 - 53)

INTEGRAL Likelihood = -0.96 Transmembrane 2 - 18 (1 - 18)

----- Final Results -----

bacterial membrane --- Certainty=0.5734(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
antigens for vaccines or diagnostics.

Example 2802

A DNA sequence (GASx1516R) was identified in *S.pyogenes* <SEQ ID 8103> which encodes the amino
acid sequence <SEQ ID 8104>. Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have no N-terminal signal sequence

-2812-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2729(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA96472 GB:AB036428 Dpr [Streptococcus mutans]
 Identities = 132/175 (75%), Positives = 153/175 (87%)

10

Query: 1 MTNTLVENIYASVTHNISKKEASKNEKTKAVLNQAVADLSVAASIVHQVHWYMRGPGFLY 60
 MINT+ ENIYAS+ H + KKE S NEKTKAVLNQAVADLS AASIVHQVHWYMRG GFLY
 Sbjct: 1 MTNTITENIYASIIHQVEKKENSGNEKTKAVLNQAVADLSKAASIVHQVHWYMRGSGFLY 60

15

Query: 61 LHPKMDLDELDSLNLNDEMSERLITIGGAPYSTLAEFKSKSLDEAKGTVDKTVQHLAR 120
 LHPKMDL+D+LN +LDE+SERLITIGGAP+STL EF ++S+L+E GT+DK++ HL R
 Sbjct: 61 LHPKMDLMDALNGHLDEISERLITIGGAPFSTLKEFDENSRLEETVGTWDKSITDHLKR 120

20

Query: 121 LVEVYLYLSSLYQVGLDITDEEGDAGTNDLFTAATEAKETIWMQLAERGQGPAL 175
 LV+VY YLSSLYQVGLD+TDEE DA +ND+FTAA+TEA+KTIWMLQAE GQ P L
 Sbjct: 121 LVQVYDYLSSLYQVGLDVTDEEDDAVSNDIFTAAQTEAQKTIWMLQAEGLQAPGL 175

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 **Example 2803**

A DNA sequence (GASx1517) was identified in *S.pyogenes* <SEQ ID 8105> which encodes the amino acid sequence <SEQ ID 8106>. Analysis of this protein sequence reveals the following:

Possible site: 46

30

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -6.32	Transmembrane	109 - 125 (106 - 126)
INTEGRAL	Likelihood = -5.26	Transmembrane	63 - 79 (61 - 81)
INTEGRAL	Likelihood = -5.20	Transmembrane	154 - 170 (151 - 176)
INTEGRAL	Likelihood = -4.14	Transmembrane	189 - 205 (189 - 205)
INTEGRAL	Likelihood = -3.50	Transmembrane	130 - 146 (127 - 147)
INTEGRAL	Likelihood = -2.92	Transmembrane	6 - 22 (1 - 24)
INTEGRAL	Likelihood = -2.23	Transmembrane	83 - 99 (83 - 101)

35

----- Final Results -----

40

bacterial membrane --- Certainty=0.3527(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

45 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA96471 GB:AB036428 type IV prepilin peptidase homologue
 [Streptococcus mutans]
 Identities = 55/127 (43%), Positives = 78/127 (61%), Gaps = 3/127 (2%)

50

Query: 83 VSASYCYLLLFSLFLDWRSQEYPPFILWLFVSFVSLLLFYISINYLSLILLLGLLAHLR 142
 ++ S LL +L SL+D + Q YP LW+ L+ Y +N +SLIL L G+ A L+
 Sbjct: 91 LTTSQVCLLFMGVLLSLYDLQDSYPLTLWIGFTFLLMFIYPLNLISLILFLFGIFAALK 150

55

Query: 143 PFSIGAGDFFYLASLALVLDLTSLIWLIQLASLAGITACLLGRIKIP--FIPYLSFGLF 200
 +IG+GDDFFYLA+LAL L+L +IW+IQ+ASL GI LL + P F+P+L G
 Sbjct: 151 NINIGSGDFFYLATLALSLNLQQTIIWIIQIASLLGILYSLLFQKHKEPPFAFVPPFLFG-H 209

Query: 201 WIVLLEH 207

-2813-

I++ H
 Sbjct: 210 LIIIFSH 216

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 5 antigens for vaccines or diagnostics.

Example 2804

A DNA sequence (GASx1538R) was identified in *S.pyogenes* <SEQ ID 8107> which encodes the amino
 acid sequence <SEQ ID 8108>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1186(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 20 antigens for vaccines or diagnostics.

Example 2805

A DNA sequence (GASx1539R) was identified in *S.pyogenes* <SEQ ID 8109> which encodes the amino
 acid sequence <SEQ ID 8110>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood =-11.73 Transmembrane 6 - 22 (3 - 32)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF31453 GB:AF221126 putative histidine kinase [Streptococcus pneumoniae]
 Identities = 141/301 (46%), Positives = 210/301 (68%), Gaps = 7/301 (2%)

Query: 1 MKRYPLLVLISYVFVIVIALITTLGLLYYTSSRNIRQLIERDTRQSIROSSQFIDAYI 60
 MKR LLV+++ +F++ + L+ +G YYQ+SS I IE +++ +I Q+S FI +YI
 Sbjct: 1 MKRSSLLVRMVISIFLVFLILLALVGTFYYQSSSAIEATIEGNSQTTISQTSHFQSYI 60

Query: 61 KPLKETTSVLAKNTEIQAFASQIHQENDKQVLQMKMVLATNSDLQAAVLVTKDGRVST 120
 K L+ T++ L+ T++ A+A Q+ + + L +L ++ DL+ VLVTK G+ +ST
 Sbjct: 61 KKLETTSTGLTQQTDVLAYAEENPSQDKVEGIRDLFLTLKSDKDLKTVVLVTKSGQVIST 120

Query: 121 NSQLTMKTSSDMMAEWPYKAAIDRQAMPILTPARQLSLSSKKEWVSVTQEVVDRAGHNL 180
 + + MKTSSDMMAE WY+ AI + AMP+LTPAR+ S +WV+SVTQE+VD G NL
 Sbjct: 121 DDSQVMKTSSDMMAEWYQKAIHQACAMPVLTTPARK----SDSQWVSVTQELVDAKANL 176

Query: 181 GVLRLDIAPYPTIKASLDQLQGRQGFAFIVNDKHEFVYHPKKSQVYSSSKEMAAMPYLA 240
 GVLRLDI+Y T++A L+QLQLG+QGFAFI+N+ HEFVYHP+ +VYSSS +M AMKPY+

-2814-

Sbjct: 177 GVLRLDISYETLEAYLNQLQLGQQGFAFTINENHEFVYHPQHTVYSSSSKMEAMKPYIDT 236

Query: 241 QNGYTKDKTSFVYQKLIPNSQWTLVGVASLDQLHRVQRQIFWSFSWNRASTLSDLWLCNCL 301

 GYT S+V Q+ I + WT++GV+SL++L +V+ Q+ W+ ++++ L +C CL

5 Sbjct: 237 GQGYTPGHKSYVSQEKIAGTDWTVLGVSSLEKLDQVRSQLLWTL---LGASVTSLLVCLCL 294

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2806

10 A DNA sequence (GASx1542R) was identified in *S.pyogenes* <SEQ ID 8111> which encodes the amino acid sequence <SEQ ID 8112>. Analysis of this protein sequence reveals the following:

Possible site: 39

>>> May be a lipoprotein

----- Final Results -----

 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC23101 GB:U32823 conserved hypothetical protein [Haemophilus influenzae Rd]
Identities = 56/128 (43%), Positives = 87/128 (67%)

Query: 73 DFELKGIDGKTYRLSEFKGKKVYLKFWASWCSICLSTLADTEDLAKMSDKDYVVLTVVSP 132
 D +LK ++ + LS++KGK VY+K WASWC ICL+ LA+ +DL+ D+++ V+T+VSP
Sbjct: 24 DVQLKDLNNQPVTL SQYKKGKPVYVKMWASWCPICLAGLAEIDDL SAEKDRNFEVITIVSP 83

Query: 133 GHQGEKSEADFKKWFQGTDYKDLPLVLLDPDGKLL EAYGVR SYPTVEVFIGSDGVLAKKHIG 192
 H+GEK ADF +W++G +YK++ VLLD G++++ VR YP +F+ SD L K G
Sbjct: 84 DHKGEKDTADFI EWYK GLEYKNITVLLDEKGEI IDKARV RGYPFNLFLDSDLNLKKT VPG 143

Query: 193 YAKKSDIK 200
 + I+
Sbjct: 144 HLGAEQIR 151

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2807

A DNA sequence (GASx1543R) was identified in *S.pyogenes* <SEQ ID 8113> which encodes the amino acid sequence <SEQ ID 8114>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -7.75	Transmembrane	171 - 187 (169 - 191)
INTEGRAL	Likelihood = -6.26	Transmembrane	205 - 221 (203 - 232)
INTEGRAL	Likelihood = -5.73	Transmembrane	56 - 72 (54 - 81)
INTEGRAL	Likelihood = -5.36	Transmembrane	92 - 108 (91 - 113)
INTEGRAL	Likelihood = -3.45	Transmembrane	20 - 36 (14 - 39)
INTEGRAL	Likelihood = -1.17	Transmembrane	147 - 163 (144 - 163)

----- Final Results -----

 bacterial membrane --- Certainty=0.4100 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-2815-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

5  >GP:AAC23102 GB:U32823 cytochrome C-type biogenesis protein
    [Haemophilus influenzae Rd]
    Identities = 106/224 (47%), Positives = 138/224 (61%), Gaps = 16/224 (7%)

Query: 6  VLMVSVFGAGLLSFFSPCIFPVLPVYLGILLDADDSKTITIFGKKLYWYGIVKTLAFIFG 65
      +L+ +VF AGL SF SPCIFP++P+Y GIL          GKK    ++ T FI G
10  Sbjct: 6  LLIGTVFLAGLASFLSPCIFPIIPIYFGILSKG-----GKK-----VLNTFLFILG 51

Query: 66  LSTIFVILGYGAGFLGNILYAVWFRYLLGALVIILGIHQMLITIKSLQFQKSLTFHNNK 125
      LS  FV LG+ GFLGNIL++  R + G +VILGIHQ+G+  I L+ K +  +
15  Sbjct: 52  LSLTFVSLGFSFGFLGNILFSNTTRIIAGVIVIVILGIHQLGIFKIGLLERTKLVEIKTSG 111

Query: 126  NRNGLFNAFILGLTFSFGWTPCVGPVLSVLALVASGGNGAWQGGVLMIIYTLGLGIPFL 185
      L  AF+LGLTFS GWTPC+GP+L+SVLAL  G+ A G  +M +Y LGL PF+
20  Sbjct: 112 KSTAL-EAFVLGLTFSLGWTPCIGPILASVLALSGDEGS-ALYGASMMFVYVLGLATPFV 169

Query: 186  LISFASGIVLKQFNKLKPHILLKKVGGVLIIVMGILLMTGTLN 229
      L SF S +LK+  L H+  K GG+LIIVMGILL+T  +
25  Sbjct: 170 LFSFFSDSLKRAKGLNKHLDKFKIGGGILIIIVMGILLITNNFS 213

```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 25 antigens for vaccines or diagnostics.

Example 2808

A DNA sequence (GASx1544) was identified in *S.pyogenes* <SEQ ID 8115> which encodes the amino acid
 sequence <SEQ ID 8116>. Analysis of this protein sequence reveals the following:

```

30  Possible site: 25

    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
35      bacterial cytoplasm --- Certainty=0.1493(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

40 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 antigens for vaccines or diagnostics.

Example 2809

A DNA sequence (GASx1546R) was identified in *S.pyogenes* <SEQ ID 8117> which encodes the amino
 acid sequence <SEQ ID 8118>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 46

    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
50      bacterial cytoplasm --- Certainty=0.4658(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-2816-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04061 GB:AP001508 unknown conserved protein in others
[Bacillus halodurans]
Identities = 48/89 (53%), Positives = 61/89 (67%)

Query: 1 MMVLVTYDVNTETPAGRKRLRHVAKLCVDYQQRVQNSVFECSTPAEFVDIKHRLTQIID 60
      M+VL+TYDV T + G KRLR VAK C +YGQRVQNSVFEC V + +K LT +ID
Sbjct: 1 MLVLITYDVQTSSMGGTKRLRKVAKACQNYGQRVQNSVFECIVDSTQLTSLKLELTSLID 60

Query: 61 EKTDSIRFYLLGKNWQRRVETLGRSDSYD 89
      E+ DS+R Y LG N++ +VE +G S D
Sbjct: 61 EEKDSLRIYRLGNKYKTKVEHIGAKPSID 89
```

- 15 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2810

A DNA sequence (GASx1547R) was identified in *S.pyogenes* <SEQ ID 8119> which encodes the amino acid sequence <SEQ ID 8120>. Analysis of this protein sequence reveals the following:

```
Possible site: 57

>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -1.70      Transmembrane      44 - 60 ( 43 - 60)

----- Final Results -----
      bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 330-332
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04060 GB:AP001508 unknown conserved protein in others
[Bacillus halodurans]
Identities = 162/341 (47%), Positives = 231/341 (67%), Gaps = 1/341 (0%)

Query: 1 MKKLLNTLYLTQEDFYVTKEGDNIVIKQEGKVLKRFPFRIIDGIVCFSTYLGVSALVKLC 60
      MKKLLNTLY+TQ D Y++ +GDN+V+ +E + L R P ++ IV F Y G S AL+ C
Sbjct: 1 MKKLLNTLYVTQPDYTLSLDGDNVLLKEQEKLGRLPLHNLEAIVGFGYT>FEATURESALMGYC 60

Query: 61 TENQINLSFHPTQGRFCGRYIGSTNGNVLLRREHYRLSDRE-ESLEYAKRFILAKISNSR 119
      E I+++F T GRF R +G + GNV+LR+ YR+S+ + ES + A+ FI K+ NS+
Sbjct: 61 AERNISITFLTKNGRFLARVVGESRGNVVLRKTQYRISENDQESTKIARNFITGKVVNSK 120

Query: 120 KYLLRFKRDRHQIDITKLFQAVNDELIWALEMVQAADNKDSLRLGIEGQAANQYFRIFNDL 179
      L R R+H +++ + F+A + L ++ ++ D+ +SLRG EGQAA Y ++F+ +
Sbjct: 121 WMLERMTREHPLRVNVEQFKATSQLSVMMQEIARNCDLSLESLRGWEGQAANYNKVFQDM 180

Query: 180 VLTDKKTFFYFQGRSKRPPLDCVNALLSFGYSLITFECQSALEAVGLDSYVGFHTDRPGR 239
      +L K+ F F GRS+RPP D VNA+LSF Y+LL + +ALE VGLD+YVGF H DRPGR
Sbjct: 181 ILQKKEEFAPFHGRSRPPKDNVNAMLSFAYTLLANDVAAALETVGLDAYVGFHMDRQDRPGR 240

Query: 240 ASLALDLVEEFRSYIVDRFVFSLINKQLQKHFVKENGSIILLTENGRAIFIDLWQKRK 299
      ASLALDL+EE R DRFV SLIN+ ++ F KENG++L+T+ R F+ WQ +K
Sbjct: 241 ASLALDLMEELRGLYADRFVLSLINRKEMTADGFYKENGAVLMTDEARKTFLKAWQTKK 300

Query: 300 HTEVEHPFTKEKVKMLLPYVQAQLLAKAIRGDLESYPPFM 340
      ++ HP+ EK+ L+PYVQA LLA+ +RGDL+ YPPF+
Sbjct: 301 QEKITHPYLGEKMSWGLVPYVQALLARFLRGDLDEYPPFL 341
```

-2817-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2811

- 5 A DNA sequence (GASx1548R) was identified in *S.pyogenes* <SEQ ID 8121> which encodes the amino acid sequence <SEQ ID 8122>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2247(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04059 GB:AP001508 unknown [Bacillus halodurans]

Identities = 90/169 (53%), Positives = 111/169 (65%), Gaps = 1/169 (0%)

Query: 45 LHTKADNPYIKEKRKELLVSRAMPISSAELGLSGIMDVVEFYKDDQGVSLRGKRGKWLPK 104
+H KAD P++KEKR L RAMPI S L +SGI DVVEF +D +G+ L G G +
Sbjct: 1 MHKADQPFMKEKRGSKLTVRAMPIQSKNLQISGICDVVEFVQDSEGIELSGVSGSYKAF 60

Query: 105 VVEYKRGKPKKDTTRDIVQLVAQTMCLEETLDCDINEGCLYYHSVNQRVIVPMTSALRQEV 164
VEYKRGKPKK DIVQLVAQ MCLEE L C I++G L+Y+ + RV VP+T ALR +V
Sbjct: 61 PVEYKRGKPKKGDEDIVQLVAQAMCLEEMLVCRIDKGYLFYNEIKHRVEVPITDALRDKV 120

Query: 165 KELAAEMHEVYQSQMLPKAAYFKNCQLCSLVDICKPRLSKKTRSVSRYI 213
++A EMH Y+++ PK C CSL IC P+L K RSV RYI
Sbjct: 121 VQMAKEMHHYENRHTPKVKTGPPCNCNLSQICLPKLMNK-RSVKRYI 168

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 Example 2812

A DNA sequence (GASx1549R) was identified in *S.pyogenes* <SEQ ID 8123> which encodes the amino acid sequence <SEQ ID 8124>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1399(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04058 GB:AP001508 unknown conserved protein in others
[Bacillus halodurans]

Identities = 148/290 (51%), Positives = 190/290 (65%), Gaps = 19/290 (6%)

Query: 6 MLEHKIDFMVTLEVKEANANGDPLNGNMPRTDAKGYGVMSDVSIKRRKIRNRLQDMGKSIF 65

-2818-

+L+HKIDF V L V +AN NGDPLNGN PR + G+G +SDV+IKRKIRNRL DM + IF
 Sbjct: 3 ILDHKIDFAVILSVTKANPNGDPLNGNRPNQYDGHGEISDVAIKRKIRNRLDMEPIF 62
 Query: 66 VQANERIEDDFRSLEKRFSQH----FTAKTPDKEIEEKANAL---WFDVRAFGQVFTYLK 118
 VQ+++R D F+SL R + K + ++E A W DVR+FGQVF +
 Sbjct: 63 VQSDDRKADSFKSLRDRADSNPELAKMLKAKNASVDEFAKACQEWMDVRSFGQVFAFKG 122
 Query: 119 K--SIGVRGPVSISMAKSLEPIVISSLQITRSTNGMEAKNNSGRSSDTMGTKHFVDYGVY 176
 S+GVRGPVSI A S++PI I S QIT+S N + RSSDTMG KH VD+GVY
 Sbjct: 123 SNLSVGRGPVSIHTATSIDPIDIVSTQITKSVNSVTGDK---RSSDTMGMKHRVDFGVY 179
 Query: 177 VLKGSINAYFAEKTGFSQEDAEAIKEVLVSLFENDASSARPEGSMRVCEVFWFTHSSKLG 236
 V KGSIN AEKTGF+ EDAE IK L++LFEND+SSARP+GSM V +V+W+ HSSKLG
 Sbjct: 180 VFKGSINTQLAEKTGFTNEDAEEKIKRALITLFENDSSSARPDGSMEVHKVYVWEHSSKLG 239
 Query: 237 NVSSARVFDLLEYHQSIIEEKSTYDAYQIHLNQEKLAKYEAKGLTLEILEG 286
 SSA+V L+ + ++D Y + L YE GL +E+++G
 Sbjct: 240 QYSSAKVHRSLKIESKTDTPKSFDDYAVEL-----YELDGLGVEVIDG 282

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2813

A DNA sequence (GASx1550R) was identified in *S.pyogenes* <SEQ ID 8125> which encodes the amino acid sequence <SEQ ID 8126>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2882(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04057 GB:AP001508 unknown [Bacillus halodurans]
 Identities = 176/671 (26%), Positives = 311/671 (46%), Gaps = 87/671 (12%)

Query: 1 MDFFTSLLKTYEKAELADLVHDQKR--NNEPVLLPIYHTSLKSNGKNIIISVKLDKDGQFH 58
 M + L +TYE A L + K+ + E LLPI HT+ ++ I V LD+DG F
 Sbjct: 1 MSWLHLHYETYE-ANLDQVGKTVKKGEDREYTLPLISHTTQNAH---IEVTLDEDGDGFL 55

Query: 59 KAEFMADKQMIIFPVTAADSVARSQSHAPHPPLVDKFAYYSAEM---GQIQ-----YDS 108
 +A+ + K+ + P T ++ +RSGS AP+PL DK +Y + + G+I+ +D+
 Sbjct: 56 RAKALT-KESTLIPCTEEAASRSGSKVAPYPLHDKLSYVAGDFVKYGGKIKNQDDAPFDT 114

Query: 109 FHKQLNNWID--YCEEVDVKKFLTFVQQFILKPEFLTLILDSLIGPDYQHNQLKVTFCD 166
 + K L W + Y E VK T++++ L + + + L NQ + +
 Sbjct: 115 YIKNLGEWANSPYATE-KVKCIYTYLKKGRLLIEDLVDAVLKL----DENQQLIEKWEK 168

Query: 167 TGKEKLIDLSACFLFESIDQ-----FQGFKNESVSTF---KALHQSYISFVEANRENLG 217
 +E L + A F + DQ F F ES+ K + S+ISF
 Sbjct: 169 RYEELLGKPAIFSSGATDQASAFVRFNVFHPESIDDVWVKDEMFDSFISFYNDKLGED 228

Query: 218 ICNISGREQITDKH---RGLMGNAKIISVS-NKREAYKGRFREREDVFSVGYETSEKI 272
 IC ++G T++H R AK+IS + N ++GRF+ + + YE S+K
 Sbjct: 229 ICFVTGNRLPSTERHANKIRHAADKAKLISANDNSGFTFRGRFKTSREAVGISYEVSKQA 288

Query: 273 HLMLKYLLENKNTSTWLGSQYILNWFSDD-LTND SRLDIVSPIFDDGLEEDDDDDTTPPV 331
 H LK+L+ ++ S + + W +D+ L + D V + E + D DT +
 Sbjct: 289 HNALKWLIRQSKSI---DDRVLVWSNDNSLVPNPDEDAVDIMKHANRELERDPDTGQI 345

-2819-

Query: 332 ITLATEDNKRIGKSFIFGQKLFANDATY---YVAILNKTSNGRIALKYFRQLQASQLLT 387
 A E K I G + +D Y ++ +L+ + GR+A+ Y+R L L
 Sbjct: 346 F--AGEVKAIGG-----YRSDLNYQPEVHILVLDSATTGRMAVLRYRSLNKELYLN 395

Query: 388 NLNKWQETYSWESRSKFGKSRLRT---PTFHDILNVSYGVDRDRFLELDNDNFKSDQIQ 443
 L W ++ +WE R + + + P DI +YG ++ D ++
 Sbjct: 396 RLEAWHDSCAWEHRYRRDEKEFISFYGAPATKDIAFAAYGPRA-----SEKVIKDLME 448

Query: 444 KLVASLIDGKPMQSIIVKKL---GNNVKEHRYRKHWYQVEQVCLAILHK---QNGEEFS 497
 +++ ++DG+ +P+ IV+ +N R+ W + + A++ K + EE+
 Sbjct: 449 RMLPCIVDGRVRPKDIVRSAFQSRASNPVSMERWE--WEKTLSTTCALIRKMHIEQKEEWG 506

Query: 498 PMLDHTNQNRSYLFGRLLAIFELIETLRYGLDGNNDNRITNAERYWTAYTGQPTKLMMLL 557
 LD ++ +RSYLFGRLLA+ +++E G G + R TNA RY +Y+ P + +
 Sbjct: 507 VPLDKSSTDRSYLFGRLLAADVLER--GALGKDETRATNAIRYMNSYKPNPGRWTWTKI 563

Query: 558 ENKIKPYEEPLKLNRRGSWMKLEKEKEEILELLNPLLETETMEKPLDYRFIFGYAEKNY 617
 + ++PY+ KL + ++ L K +EI + P + PL +++ G+Y+++
 Sbjct: 564 QESLQPYQ--AKLGTKATY--LSKLVDEIGDQFEP--GDFNNNPLTEQYLLGFYSQRRE 616

Query: 618 YYTKQNTTEVTE 628
 Y K+ E +
 Sbjct: 617 LYKKKEEETNQ 627

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2814

A DNA sequence (GASx1551R) was identified in *S.pyogenes* <SEQ ID 8127> which encodes the amino acid sequence <SEQ ID 8128>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3035(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04056 GB:AP001508 unknown [Bacillus halodurans]

Identities = 90/218 (41%), Positives = 127/218 (57%), Gaps = 7/218 (3%)

Query: 13 GQRALFTNPATKGGSSRSYSVPTRQALNGIVDAIYYKPTFTNIVTEVKVINQIQTELQG 72
 G ALFT+P TK G E+ SYSVPT QAL GI ++IY+KPT ++ E++V+ IQ E +G
 Sbjct: 11 GDYALFTDPLTKIGGEKLSYSVPTQALKGIAESIYWKPTIVFVIDELRVMKPIQMESKG 70

Query: 73 VRALLHDYSADLSYVSYLSDVVYLIKHFVWNEDRKDLNSDRLPAKHEAIMERSIRKGGGR 132
 VR + + L++ +YL DV Y +K HF +N R DL DR KH +I++RS++ GGR
 Sbjct: 71 VRPIEYGGGNTLAHYTYLKDVBHYQVKAHFEFNLHRPDLAFDRNEGKHYSILQSLKAGGR 130

Query: 133 RDVFLGTRECLGLVDDISQEEYETTYSYNGV-NIDLGIMFHSFAYPKDK-KTPLKSYFT 190
 RD+FLG REC G V + E+ + +Y+G LG M H F YP + + L
 Sbjct: 131 RDIFLGARECQGYV--APCEFGSGDGFYDGGKYHLGTMVHGFPNYPDETGHQQLDVRWLW 187

Query: 191 KTVMKNGVITFKAQSECDIVNTLSSYAFKA--PEEIKS 226
 VM+NG I F +C IV + K P+ ++S
 Sbjct: 188 SAVMENGYIQFPRPEDCPVRPVKEMEPKIFNPDNVQS 225

-2820-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2815

A DNA sequence (GASx1552R) was identified in *S.pyogenes* <SEQ ID 8129> which encodes the amino acid sequence <SEQ ID 8130>. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2770(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04055 GB:AP001508 unknown conserved protein in others
[Bacillus halodurans]

Identities = 252/836 (30%), Positives = 404/836 (48%), Gaps = 90/836 (10%)

Query: 3 MILAHYDCKKDKKQSLDEHLWHVACSSRQEASIIQGQDVLFLIGLYHDLGKADRTFQD-- 60
M +AH Q+L EHL V C + + + V L GL HDLGK F+D
Sbjct: 1 MYIAHIREVDKVIQTLKEHLGCGVQCLAETFGAKLRLQHVAGLAGLLHDLGKYTNEFKDYI 60

Query: 61 -----KLLNNPNRHHVDHSYAGAKYLCIIIPHLKNRGVDKNERMTFNEMVGIVISAHH 113
+L VDHS AG + L + L +R +E++ E+VG I +HH
Sbjct: 61 YKAVFEPPELAEKKRGQVDHSTAGGRLLYQM----LHDRENSFHEKL-LAEVVGNNAIISHH 115

Query: 114 GMYDLCYYFDDAEYYGFNKFKNRINRDLGYYHYHEDIKGYALKLEKKLCDYGYK-DLREL 172
+Y N + R L+ +++ Y +E+ + + +L
Sbjct: 116 SNLQ-----DYISPTIESNFLTRVLE-----KELPEYESAVERFFQEVMTAEALARY 162

Query: 173 IDKAFDNYQQAMSSSLNWQDKSEWYYQSCMVRLYLSLLKNADILDITVNAVGLKISPMDKT 232
+ KA D +Q + Q Y SC++ +AD +T + + + T
Sbjct: 163 VAKAVDEIKQFTDNSPTQSFFLTQYIFSLI-----DADRTNT-RMFDEQAREEEPT 213

Query: 233 ERSFLKHSYLAIEQKYASFGQPNNQ---LNTIRTEIAERVKERGKRDSKGIYRLDLPTG 289
+ L Y + AS + ++ +N +R+ ++E+ + R S GIY L +PTG
Sbjct: 214 QPQQLFEBHYHQQLLNHLASLKESDSAQKPINVLRMSAMSEQCESFAMRPS-GIYTLSIPTG 272

Query: 290 AGKTNLSMRYAFHQVLVHDKSRFFYITPFLSVLEQNASEIRKVTGD-LGVLEHHSNVVKQ 348
GKT S+RYA ++K R YI PF +++EQNA E+R + GD +LEHHSNVV+
Sbjct: 273 GGKTLASLRYALKHAQEYNKQRIIYIVPFTTIEQNAQEVNRNILDENILEHHSNVVED 332

Query: 349 ANEDDDDKDSLISA-----YLSDSWDSQVVLTSMVQFFQTLFKTKSANLRFRSSSLINSV 403
+ D+ +D +++ D+WD ++ T++VQF + + N RR +L +SV+
Sbjct: 333 SENGDEQEDGVITKKERLRLARDNWRPIIFTTLVQFLNVFYAKGNRNRRLHNLHSHSVL 392

Query: 404 ILDEVQSLPIEVTTLEFNLTMNFLNKVMDTTIVLCTATQPAYDSSEIDHRICYGGLGELA 463
I DEVQ +P + +LFN +NFL + +I+LCTATQP ++ + H + +
Sbjct: 393 IFDEVQKVPTKCVSLFNEALNFKFAHCSILLCTATQPTLEN--VKHSLKDRD---G 446

Query: 464 EIVELTIEEKQIFSRTELRFDDSDQKVHLTDVINLILGEE---NSVLAIFNTKKTIVHNC 520
EIV+ E + F R E+ D +DQ + + + E S L I NTKK V +
Sbjct: 447 EIVQNLTEVSEAFKRVEI--LDKTDQPMTNERLAEWVRDEAPSWGSTLIILNTKKVVKDL 504

Query: 521 YTMLKMDTRPVYQLSTNMCAQHRLDLIAKIKTELQNNIPIICISTQLIEAGVDVDFHRV 580
Y L+ PV+ LST+MCA HR D + +I+ L+ P IC++TQLIEAGVDV F V
Sbjct: 505 YEKLEG-GPLPVFHLSTSMCAHRKQQLDEIRALLKEGTPFICVTTQLIEAGVDVSFKCV 563

Query: 581 IRSYSGIDSIVQAAGRCNREGKRDKGQVTLVNLNTEENISRLTEIKTKKEATESILHKI 640

-2821-

```

      IRS +G+DSI QAAGRCNR G+      V +++ + EE +S+L EI+ +E ++L +
Sbjct: 564 IRSLAGLDSIAQAAGRCNRHGEEQLQYVYVID--HAETLSKLKEIEVGQEIAGNVLARF 621

Query: 641 GSPIDISTLN-----RDFFEYYYANNQGLMDYPLED-----NLSTYDYLSTLNIYTAN 688
      +   N      R++F YYY+      ++Y +++      +   +   N Y T
Sbjct: 622 KKAKEKYEGNLLSQAMREYFRYYYSKMDANLNYFVKEVDKDMTKLLMSHAVENSYVTTY 681

Query: 689 KKFQKG-----LKQAFKTAGAKMNLINNDMIGILVPYGEAEKKLAYLEELGVSHFLSAKD 743
      +K G      L ++KTA      +I+ +      +VPYGE + +A L      S +
Sbjct: 682 QKNTGTHFPLLLLNGSYKTAADHFRVIDQNITSAIVPYGEGQDIIAQLN-----SGEW 733

Query: 744 YQTIKSLLELQPFITVNV--RENDPLFE--TTKSYLNGQILVLTSEYYDTERGVKY 795
      + +LK+ Q +TVN+ +E D L +      +L+G + L +Y + GV +
Sbjct: 734 VDDLKVLKKAQQYTVNLYSQEIDQLKKEGAIVMHLDMVYELKESWYSHQYGVDF 789

```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2816

A DNA sequence (GASx1558) was identified in *S.pyogenes* <SEQ ID 8131> which encodes the amino acid sequence <SEQ ID 8132>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

      bacterial cytoplasm --- Certainty=0.1050(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2817

A DNA sequence (GASx1563) was identified in *S.pyogenes* <SEQ ID 8133> which encodes the amino acid sequence <SEQ ID 8134>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

      bacterial cytoplasm --- Certainty=0.1872(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2818

A DNA sequence (GASx1564R) was identified in *S.pyogenes* <SEQ ID 8135> which encodes the amino acid sequence <SEQ ID 8136>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 32

   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2173 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

15 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2819

A DNA sequence (GASx1566R) was identified in *S.pyogenes* <SEQ ID 8137> which encodes the amino acid sequence <SEQ ID 8138>. Analysis of this protein sequence reveals the following:

```

20   Possible site: 43

   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
25      bacterial cytoplasm --- Certainty=0.3486 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

30 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2820

A DNA sequence (GASx1568) was identified in *S.pyogenes* <SEQ ID 8139> which encodes the amino acid sequence <SEQ ID 8140>. Analysis of this protein sequence reveals the following:

```

35   Possible site: 26

   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
40      bacterial cytoplasm --- Certainty=0.2711 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

45 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2823-

Example 2821

A DNA sequence (GASx1569) was identified in *S.pyogenes* <SEQ ID 8141> which encodes the amino acid sequence <SEQ ID 8142>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2822

A DNA sequence (GASx1576R) was identified in *S.pyogenes* <SEQ ID 8143> which encodes the amino acid sequence <SEQ ID 8144>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4042 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2823

A DNA sequence (GASx1577R) was identified in *S.pyogenes* <SEQ ID 8145> which encodes the amino acid sequence <SEQ ID 8146>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3342 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04515 GB:AP001509 unknown [Bacillus halodurans]
Identities = 36/104 (34%), Positives = 55/104 (52%)

-2824-

Query: 2 HMGAWNTGNNKILYTQESVTDMMIAKRDQSIKDAKESPILGFTVDTKVIKTELSNISNVM 61
 +M ++ GN IL E D + + A SP LGF D+ ++TE++ ISNV
 Sbjct: 392 NMPSFAIGNQLILKLYEDDPQDKWEAFEFNESAIPSPALGFYFDSNPVTEIAAISNVT 451

Query: 62 NRYKASINTGTVPDEALPKLLADLKGAGWDKVQKEVQKQLDDF 105
 + + ++ G VDP+E LP L AG KV E+Q+Q D++
 Sbjct: 452 SEFSPALLKGAVDPPEEYLPLEFNDKLEAGLQKVIDEMQRQFDEW 495

- 10 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2824

A DNA sequence (GASx1578R) was identified in *S.pyogenes* <SEQ ID 8147> which encodes the amino acid sequence <SEQ ID 8148>. Analysis of this protein sequence reveals the following:

15 Possible site: 27

>>> May be a lipoprotein

----- Final Results -----

20 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

- 25 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04515 GB:AP001509 unknown [Bacillus halodurans]
 Identities = 134/346 (38%), Positives = 206/346 (58%), Gaps = 10/346 (2%)

Query: 21 AACESKASKDSV KLLMYQVGDKPDNFDELMTIANKRIKEKTGATVDLQYIGWGDWDDK 80
 +A E+++ D V L Y +G + + +M N +EK ATVDL+ + WG++D++
 Sbjct: 42 SANETEATDLDH-VTLTWYMGITPQPDLELVMEEVNAYTEEKINATVDLRMLDWGEYDER 100

Query: 81 MSTIIASGENYDIAF----ANNYVVNAQKGAFADLTTLMPKYAKKTYKNLDPAYIKGNTI 136
 M I SGE YDIAF ANNY +NA++GAF +L L+ ++ ++ + +DPA+++G +
 Sbjct: 101 MQVITTSGEAYDIAFTSSWANNYALNARRGAFLELNDLLDEHGQEMKELIDPAFLEGAQV 160

Query: 137 DGKLYA FPDANVYAQQMLSFNKELVDKYGLDISNIKSYADAENVLKQFHEKEPNTAAFA 196
 DGKLYA P + V Q +LSFN ELV+K+ LD+S++ S AD E +L E+E + A
 Sbjct: 161 DGKLYAVPTNKEVGQQA VLSFNELVEKHNLDLSSVHSLADLEPLLAVIKEEESDVTPIA 220

Query: 197 IGQVFSMSGDYDYPLTKTQPFQAVKIDEGKPTIINQYEDSFKNLRLMHKWKYKEGLIPTD 256
 F +D L + PFA +++ +IN+YE++ L+ MH +YK+G I D
 Sbjct: 221 ---TFDAYLPFDSILQEEMPFARLEGNTNEVINKYEEDITMETLKTMDHYKKGYIRPD 277

Query: 257 AATNTEGYPLEGNTWFMREETQGPMDYGDITLTNAAGKDIVSRPLTKPLKTTTSAQMANF 316
 AAT+T+ +PLE WF+R+E P Y + I T AG +I +RPL +P + +
 Sbjct: 278 AATSTDSWPLETPNWFVRKELYQP--YAELIWTRTAGYEIATRPLHEPYIFNNSVTGSMQ 335

Query: 317 VVSSVSKNKEKAVEVLSLLNSDPELLNGLVYGVEGKAWEKIGDKKI 362
 +S+ SKN E+A+ L+LLNSDP L N L G+EG +E++ D I
 Sbjct: 336 AISATSKNPERAMMFLNLLNSDPYLRNLLDKGIEGVHYEELEDGTI 381

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2825-

Example 2825

A DNA sequence (GASx1582) was identified in *S.pyogenes* <SEQ ID 8149> which encodes the amino acid sequence <SEQ ID 8150>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```
bacterial cytoplasm --- Certainty=0.0454 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2826

A DNA sequence (GASx1584R) was identified in *S.pyogenes* <SEQ ID 8151> which encodes the amino acid sequence <SEQ ID 8152>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```
bacterial cytoplasm --- Certainty=0.3105 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

RGD motif: 3-5

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG21428 GB:AF307332 meningioma-expressed antigen 5s splice
variant [Homo sapiens]

Identities = 94/271 (34%), Positives = 148/271 (53%), Gaps = 14/271 (5%)

```
Query: 120 GIIEGFYGTPTWREERLDCRFIGNKRMNTYMYAPKDDDYQRKLWRDLYPEDWVTYFKEL 179
      G++EGFYG PW E+R + R + +NTY+YAPKDD R WR++Y + L
Sbjct: 63 GVVEGFYGRPWVMEQRKELFRRLQKWLNTYLYAPKDDYKHRMFWRMYSVEEAEQLMTL 122
```

```
Query: 180 LAVAKEEGLDFWYMISPLDIFYTKEADYQLLYQKLQQLLALGVCHFGLLLDDIDYQIVD 239
      ++ A+E ++F Y ISPLD ++ + L +KL Q+ G F LL DDID+ +
Sbjct: 123 ISAAREYEIEFIYAIISPLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFDDIDHNMCA 182
```

```
Query: 240 AVERRFKKTAYAQAHATEVHHFLNQQAAPELVICPTE-----YDNHHDSIYLQELSE 293
      A + F A+AQ + E++ +L + + CPT E Y N S YL+ + E
Sbjct: 183 ADKEVFSSFAHAQVSI'TNEIYQYLGEPEP---FLFCPT EYCGTFCYPNVSQSPYLRTVGE 239
```

```
Query: 294 RIPKEVAFFWTGPSTLASQISQADIETMAAVYQRPILIIWDNIPVNDYQKDPERLFLTPFA 353
      ++ + WTGP ++ +I IE ++ + +R +IWDNI NDY D +RLFL P+
Sbjct: 240 KLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDY--DQKRLFLGPYK 297
```

```
Query: 354 NRSPFLCQPDYQVKGIVSNPMISWELSKLTL 384
      RS L ++KG+++NP +E + + +
Sbjct: 298 GRSTELIP---RLKGVLTNPCEFEANYVAI 325
```

-2826-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2827

A DNA sequence (GASx1585R) was identified in *S.pyogenes* <SEQ ID 8153> which encodes the amino acid sequence <SEQ ID 8154>. Analysis of this protein sequence reveals the following:

Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4469(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2828

A DNA sequence (GASx1587) was identified in *S.pyogenes* <SEQ ID 8155> which encodes the amino acid sequence <SEQ ID 8156>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3082(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04509 GB:AP001509 unknown conserved protein in others

[Bacillus halodurans]

Identities = 221/425 (52%), Positives = 296/425 (69%), Gaps = 4/425 (0%)

Query: 12 RPIPTSVSQFMAKVESLQHPDWALNFKTSFTNTLETTLKTYEDGTSFLLTGDI PAMW 71

+ IP S+ +A+V++ D L F+ F NT TT++ E GT F++TGDIPAMW

Sbjct: 4 KKIPRSLQAIQAQVKAHYADDQELQTL-FEQCFLNITYLTITIQEDEQGT-FVVTGDIPAMW 61

Query: 72 LRDSTAQMKPYLFLAKEDEEIRKIIAGLVKRQFRYICIDPYANAFNEEANEKGHQTDHTQ 131

LRDS+AQ++PYL + KED ++ ++I G+++RQ+RYI DPYANAFN+ AN++GHQ D T+

Sbjct: 62 LRDSSAQVRPYLTVVKEDADMARMIKGVIERQWRYILHDPYANAFNQTANKQGHQQRTE 121

Query: 132 MNPWIWERKYEIDCLCYPIQLAYLLYRETGSTDQFNDDFHRGVLEILDLWTVEQDH-AQS 190

M+P +WERKYE+D LCYPIQLAYL ++ TG + +E I +W +EQDH A+S

Sbjct: 122 MSPLVWERKYELDSL CYPIQLAYLYWKATGDDSVLQPTLQVLETIYRIWKIEQDHEAKS 181

Query: 191 PYLFERDITWRKEDTLTHAGKGSFPVPTGMTWSGFRPSDDACQYGYLIPSNMFAVVVLSYL 250

Y FERD R DTL GKG PTGMTWSGFRPSDDAC YGYLIP+NMFAVVV +Y

Sbjct: 182 SYSFERDDCRVSDTLRLKKGKGYSVPTGMTWSGFRPSDDACLYGYLIPANMFAVVVSNYA 241

Query: 251 EDLYNNLFHNEPVPATRAKQLKEAIQSGIADHALVQNSKGETIYAYEVDGLGQFSIMDDAN 310

+L + +A ++L+ I+ GI + + + IY YE DG G+ ++MDDAN

-2827-

Sbjct: 242 VELLTAM-EEIKLAEEFRELEADIRQIGIQYGMHPVYGEIYVYETDGNRNVNLMDDAN 300

Query: 311 IPSLLAAPYLGFCTKDDPIYLATRRITLSQENPYYYQGNAAAGIGSSHTPENIWHIALA 370
 +PSLLA PYLG+ T DDP+Y TRR ILS++NPYYY+G+ A G+GS HTP++Y+WHI+LA

5 Sbjct: 301 VPSLLAIPYLGYYTADDPVYQNTRRFILSRDNPYYYEGSYAKGVGSPHTPDHYVWHISLA 360

Query: 371 LQGLTALDQDSKKEMLDLLVATDAGTHLMHEGFVDNDPYQYTREWFSWANMMFCELLLDY 430
 +QG+TA+D KK+++ + T A T+ MHEGFV+ P QYTR WF+WAN MF E LL

10 Sbjct: 361 IQGMTAIDSKEKKQIVAMFKQTHADTYFMHEGFVDVRPEQYTRSWFAWANSMFSEFLLSE 420

Query: 431 LGFSI 435
 G +

Sbjct: 421 AGIYV 425

15 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2829

A DNA sequence (GASx1588) was identified in *S.pyogenes* <SEQ ID 8157> which encodes the amino acid sequence <SEQ ID 8158>. Analysis of this protein sequence reveals the following:

20 Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.5250(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

30 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04508 GB:AP001509 unknown conserved protein in others
 (divided) [Bacillus halodurans]
 Identities = 312/737 (42%), Positives = 426/737 (57%), Gaps = 21/737 (2%)

35 Query: 123 FPDTFGNGQTPQLMLKAGLQAAAFGRGIRPTGFNNQVDTSKYSSQFSEISWQGPDNSR 182
 FPDTFG GQ PQL+ +AG++AA FGRG+ PTGFNNQV + YSS FSE+ W+ PD S+
 Sbjct: 4 FPDTFGIYQAPQLLAQAGIRAAVFGRGVPTPTGFNNQVQHDD-YSSPFSELIWEAPDGSQ 62

40 Query: 183 ILGLLFANWYSNGNEIPTTEAEARLFWDKKLADAERFASTKHLMMNGCDHQPVQLDVTQ 242
 ++G+L ANWYSNGNEIPT E EA+ FW KKL DAERFAST LL MNGCDHQPVQ DVT+
 Sbjct: 63 VIGILLANWYSNGNEIPTDEEAQTFWVKKLDAERFASTSOLLFMNGCDHQPVQKDVTD 122

Query: 243 AIALANQLYPDYEFVHSCFEDYLADLADLPENLSTVQGEITSQETDGWYTLANTASARI 302
 AI +A L+PD F HS F DYL + ++LP+ L + GE+ +Q+TDGW TL NTASARI

45 Sbjct: 123 AIKVAETLFPDVAFKHSNFHDYLTQIKEELPKELQKITGELRNQKTDGWSTLVNTASARI 182

Query: 303 YLKQANTRVSRQLENITEPLAAMAYEVTSTYPHDQLRYAWKTLMNHPHDSICGCSVDSV 362
 YLKQAN R L N+ EP+ + + D Y WK LM+NHHPHDSICGCS+D+V

50 Sbjct: 183 YLKQANDRCQTLTNVLEPMCLLV--ENKSLHRDFSEYYWKLLENHHPHDSICGCSIDAV 240

Query: 363 HREMTRFEKAYEVGHYLAKEAAKQIADAIDTRDFPMDSQPFVLFNTSGHSKTSVAELSL 422
 HREM TRFEK E K+IA I+T ++ P V+ T+G S V +

Sbjct: 241 HREMTRFEKVEAGATTFAEQGKEIAAQINTLHDSEEAIPLVLKTINGTSGKRVVRHKV 300

55 Query: 423 TWKKYHFGQRFPPKEVYQEAQEYLARLSQSFIIDTSGQVRPEAEILGTSTAFDYDLPKRS 482
 KK +F + ++ + L + ++ + E+ + F YDL+P

Sbjct: 301 AMKKIYFDEM---DFRHIPDRLEKIVMPTYRLEFPNKGSVPIEVQDAGVRFGYDLPRDG 356

Query: 483 FREPYFAIKVRLRLPITLPAMSWKTLALKLG-----NETTPSETVSLYDDSNQCLENGF 536
 FR PY+A L +T S L + G + T + + D S LEN

60 Sbjct: 357 FRRPYA----RELEVTFYSYSDLYLGYECGFLVPVEEKQTEARKELIGDPSMNTLENEA 412

-2828-

Query: 537 LKVMIQTDGRLTITDKQSGLIYQDLRFEDCGDIGNEYISRQPNHDQPFYADQGTIKLNI 596
 +KVM I +G +I DK +G Y+ L +ED GDIGNEY+ + + + + I
 Sbjct: 413 MKVMIHRNGSYSILDKTTGFYRHLGIYEDVGDIGNEYMFKASSDGVRYTTEACEASIRI 472

Query: 597 ISNTAQVAELEIQQTFAIPISADKLLQAEMEAVIDITERQARRSQEKAELTLTTLIRMEK 656
 I N + A +EI QT ++P +AD+ L+ E E ++ +R+A RS+E+ ++TL T + +E+
 Sbjct: 473 IENNSLCATVEICQTLSPAAADERLKEEQERLVWHPDRKAGRSKERTDITLRTLELTLEQ 532

Query: 657 NNPRLQFTTRFDNQMTNHLRLVLPFTHLKTIDHHLADSI FETVKRPNHPDATFWKNPSNPQ 716
 L+ DN +HR+R LFP +H ADSI+E V+RPN PD W+NP+
 Sbjct: 533 GAKGLKVNVDINTAKDHRMRALFPVERARGNHYSIYIVERPNTDPDK-WQNPAFDH 591

Query: 717 HQECFVSLFDGENGVTIGNYGLNEYEILPDTNTIAITLLRSVGEMGDWGYFPTPEAQCLG 776
 H + VSL +GE G+TI GL+EYEI+ D +IA+TLRSVGE+GDWG F TPEAQC G
 Sbjct: 592 HMQLVSLDNGEYGLTIATKGLHEYEIVSD--SIAVTLRSVGELGDWGLFETPEAQCFCG 649

Query: 777 KHSLSYSFESITKQTQFAS-YWRAQEGQVPVITTTQTNQHEGTLAAEYSYLTGTNDQVALT 835
 ++ + A+ Y A + V QT Q G L + + + + LT
 Sbjct: 650 QNEAQFVLLPHKGDVLSANVYVAAYDDPVEPTVIQTEQSMGPLPHATNLFQWSGEGVLVT 709

Query: 836 AFKRLADNALITRSYN 852
 A K + +I R +N
 Sbjct: 710 ACKPTMDGRGMILRWFN 726

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2830

A DNA sequence (GASx1589R) was identified in *S.pyogenes* <SEQ ID 8159> which encodes the amino acid sequence <SEQ ID 8160>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -11.30	Transmembrane	203 - 219 (195 - 221)
INTEGRAL	Likelihood = -8.17	Transmembrane	61 - 77 (59 - 82)
INTEGRAL	Likelihood = -3.98	Transmembrane	107 - 123 (107 - 124)
INTEGRAL	Likelihood = -3.40	Transmembrane	39 - 55 (38 - 58)
INTEGRAL	Likelihood = -2.34	Transmembrane	129 - 145 (126 - 145)
INTEGRAL	Likelihood = -2.07	Transmembrane	89 - 105 (87 - 105)

----- Final Results -----

bacterial membrane	---	Certainty=0.5522(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAC10175 GB:AJ278302 histidine kinase [Streptococcus pneumoniae]
 Identities = 114/432 (26%), Positives = 219/432 (50%), Gaps = 10/432 (2%)

Query: 21 LTLKLSFVSIAIPRLKNIYFLSLSMVLFQVVFVAFPPDHFILDVVM LAQF---LFFALI 77
 L + +F V I L + IF L +L VVF +++ V L+ F L+ +
 Sbjct: 16 LKIVIFFKVDGISLTFERIFKAFLFKILLAVVFGML---GYMVGNVYLSYFMEPLYGIGL 72

Query: 78 ALYYGKSIKAKFLMFYAFFPLVSI SLVKRFIVFFVMPLFGMPYSVVKHNTLLIYSITCFS 137
 + + + K L+FY FP++ ++L R + +FV+P G V + + I F+
 Sbjct: 73 SFLLRLRELPKLLLFYGLFPMILVNLFYRGVSYFVLPFLGQG-QVYDDYSFIWLCIIIFN 131

Query: 138 IFLIYRCIQVFHFDFSTWRQYFQSHRASKLLVFTNSSMALYYLCVQGIDVMSPSLSGLAT 197
 F+ ++ +DF++ R+ K L N M YYL +Q + G+ +
 Sbjct: 132 FFISLAFLLKWLDDYDFTSLRKGILDKDFQKSLTQINWIMGAYYLVQNL SYFEYE-QGIQS 190

-2829-

Query: 198 TTARSIIVLFYFILFTLLIHLERYVKQNSIEAIVQOKE--YRELINYSQHLGLLYQDIQ 255
 TT R +I++FY + F+ ++ L+ Y+K E + Q+++ YRE+ YS+H+ LY++++
 Sbjct: 191 TTVRHLLLVFYLLFFMGIIKKLDITYLKDKLHERLNQEQDLRYREMERYSRHIEELYKEVR 250
 5
 Query: 256 ELRRLLTTVSSRLKIGIEQNDISIVRLTYEGILNAEKNNAKDDRLDLTCLDKLQVEAIRH 315
 R T + + L++GIE+ D+ ++ Y+ +L +D++ DL L ++ A++
 Sbjct: 251 SFRHDYTNLLTSLRLGIEEEDMEQIKEIYDSVLKDSSEKLQDNKYDLGRLVNVDRALKS 310
 10
 Query: 316 IVLAKLIEAKNKKLKVEVSIPNCIATFFLEVVDFTKLLSFLLDNAIEMSLETQKPCLSIA 375
 ++ K I+A++K + V +P I + ++DF ++S L DNAIE S+E QP +SIA
 Sbjct: 311 LLAGKFIKARDKNIVFNVEVPEEIQVEGVSLDFLTIVVSILCDNAIEASVEACQPHVSIA 370
 15
 Query: 376 FLDQNHKLIVIVIQSSTKQGGDSQSVFAIPALKKRDDWQFDLRNVTTILNRYDYLTISSQ 435
 F + +I++S K+ D +F+ A K ++ L V I+ + +++
 Sbjct: 371 FFKNGAQETFIENSIKEEGIDISEIFSGASSKGEERGVLGYTMKIVESHNPNTSLNTT 430
 20
 Query: 436 IHDGILTQLIEI 447
 D + Q++ +
 Sbjct: 431 CQDHVFRQVLTV 442

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2831

25 A DNA sequence (GASx1593R) was identified in *S.pyogenes* <SEQ ID 8161> which encodes the amino acid sequence <SEQ ID 8162>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq
 30 INTEGRAL Likelihood = -1.28 Transmembrane 2 - 18 (1 - 18)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 40 antigens for vaccines or diagnostics.

Example 2832

A DNA sequence (GASx1594) was identified in *S.pyogenes* <SEQ ID 8163> which encodes the amino acid sequence <SEQ ID 8164>. Analysis of this protein sequence reveals the following:

Possible site: 61
 45 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.93 Transmembrane 76 - 92 (76 - 92)
 ----- Final Results -----
 50 bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

-2830-

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF61313 GB:U96166 unknown [Streptococcus cristatus]
  Identities = 31/66 (46%), Positives = 40/66 (59%), Gaps = 2/66 (3%)

5   Query: 14 LLGRILSKYVGRLTSCIENETTKIRNHSRQNDTIGLNHLLGNLKTVHNPEIILKTINVYS 73
      + G +SK + + E K+ ++ ND IG N LLG+LKTVHNPEII + VYS
      Sbjct: 30 VFGMDVSKTSSEVAILVNGE--KVHGYTILNDAIGFNRLIGDLKTVHNPEIIFEATGVYS 87

      Query: 74 RRLQVF 79
      RRLQ F
10     Sbjct: 88 RRLQAF 93
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

15 Example 2833

A DNA sequence (GASx1598) was identified in *S.pyogenes* <SEQ ID 8165> which encodes the amino acid sequence <SEQ ID 8166>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

20  >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2117(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
30 antigens for vaccines or diagnostics.

Example 2834

A DNA sequence (GASx1608) was identified in *S.pyogenes* <SEQ ID 8167> which encodes the amino acid sequence <SEQ ID 8168>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

35  >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
40  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
45 antigens for vaccines or diagnostics.

-2831-

Example 2835

A DNA sequence (GASx1619) was identified in *S.pyogenes* <SEQ ID 8169> which encodes the amino acid sequence <SEQ ID 8170>. Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2916(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2836

A DNA sequence (GASx1621) was identified in *S.pyogenes* <SEQ ID 8171> which encodes the amino acid sequence <SEQ ID 8172>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1899(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

alpha subunit [Escherichia coli]
Identities = 110/211 (52%), Positives = 153/211 (72%)

Query: 7 KEITIKEAVAHVKDGTIMVGGFMTNGTPEKLIDALVEKGVKDLTLICNDAGFPDKGVGK 66
K +T+++A +DG TIMVGGFM GTP +L++AL+E GV+DLTLI ND F D G+G
Sbjct: 4 KLMTLQDATGFFRDGMTIMVGGFMGIGTPSRLVEALLESQVRDLTLIANDTAFVDTGIGP 63

Query: 67 MVANKQFSTIIASHIGLNREAGRQMTGETVIDLVPQGTLAERIRSGGFLGGFLTPTGI 126
++ N + +IASHIG N E GR+M GE + LVPQGT L E+IR GG GLGGFLTPTG+
Sbjct: 64 LIVNGRVRKVIASHIGTNPETGRRMISGEMDVVLVPQGTLEQIRCGGAGLGGFLTPTGV 123

Query: 127 GTEVAKGKEVITIDGKDYLLKPLKADVALIFANKADKNGNLQYAGSENNFNHVMANAK 186
GT V +GK+ +T+DGK +LLE+PL+AD+ALI A++ D GNL Y S NFN ++A A
Sbjct: 124 GTVVEEGKQTLTLDGKTWLLERPLRADLALIRAHRCDTLGNLTQLSARNFNPLIALAAD 183

Query: 187 TTIVEAREIVDVGQMDPNFVHTPGIFVNYLV 217
T+VE E+V+ G++ P+ + TPG +++++
Sbjct: 184 ITLVEPDELVETGELQPDHIVTPGAVIDHII 214

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2832-

Example 2837

A DNA sequence (GASx1622) was identified in *S.pyogenes* <SEQ ID 8173> which encodes the amino acid sequence <SEQ ID 8174>. Analysis of this protein sequence reveals the following:

Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4668(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD54948 GB:AF157306 acetoacetate:butyrate/acetate coenzyme A
transferase [Clostridium beijerinckii]
Identities = 121/214 (56%), Positives = 161/214 (74%), Gaps = 5/214 (2%)

Query: 7 VLSKEEIQTRIAKRVAQELEHNTLVNLGIGLPTKVANYIPEGVTITLQSENGFVGLTGLT 66
VL+KE I AKRVA+EL+ LVNLGIGLPT VANY+P+ + IT +SENG VG+ +
Sbjct: 6 VLAKEII----AKRVAKELKKGQLVNLGIGLPTLVANYVPKEMNITFESENGMVGMQAQMA 61

Query: 67 DD-HYDPTIVNAGGQPVSIAPGGAFFDSSTSFGIIRGGHVAATVLGALQVDKEASIANYL 125
DP I+NAGG+ V++ P GAFFDSSTS+IRGGHV VLGA+VD+E ++AN++
Sbjct: 62 SSGENDPDIINAGGEYVTLPLQGAFDSSTSFFALIRGGHVDVAVLGALEVDDEGNLANWI 121

Query: 126 IPGKMVPGMGGAMDLLVGAKKVIVAMEHTNKGKAKILDKCTLPLTAQNVVNLIIITEMGVF 185
+P K+VPGMGGAMD+GAKK+IVAM+HT KGK KI+ KCTLPLTA+ V+LI+TE+ V
Sbjct: 122 VPNKIVPGMGGAMDLAIGAKKIIVAMQHTGKGPKIVKKCTLPLTAKAQVDLIVTEL CVI 181

Query: 186 EYQDEGLCALEINPDYTFEDVQNVTEVTLLIDKTN 219
+ ++GL EI+ D T ++++ +T+ LI N
Sbjct: 182 DVTNDGLLFRFRIHKDTTIDEIKFLTADADLIIPDN 215

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2838

A DNA sequence (GASx1628R) was identified in *S.pyogenes* <SEQ ID 8175> which encodes the amino acid sequence <SEQ ID 8176>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1243(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2839

A DNA sequence (GASx1639R) was identified in *S.pyogenes* <SEQ ID 8177> which encodes the amino acid sequence <SEQ ID 8178>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -8.65	Transmembrane	55 - 71 (44 - 73)
INTEGRAL	Likelihood = -7.64	Transmembrane	13 - 29 (5 - 31)

----- Final Results -----

bacterial membrane	---	Certainty=0.4461(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2840

A DNA sequence (GASx1643) was identified in *S.pyogenes* <SEQ ID 8179> which encodes the amino acid sequence <SEQ ID 8180>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.0766(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2841

A DNA sequence (GASx1645R) was identified in *S.pyogenes* <SEQ ID 8181> which encodes the amino acid sequence <SEQ ID 8182>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside	---	Certainty=0.3000(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2834-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2842

A DNA sequence (GASx1649R) was identified in *S.pyogenes* <SEQ ID 8183> which encodes the amino acid sequence <SEQ ID 8184>. Analysis of this protein sequence reveals the following:

Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0931(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2843

A DNA sequence (GASx1650) was identified in *S.pyogenes* <SEQ ID 8185> which encodes the amino acid sequence <SEQ ID 8186>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5678(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2844

A DNA sequence (GASx1651R) was identified in *S.pyogenes* <SEQ ID 8187> which encodes the amino acid sequence <SEQ ID 8188>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2761(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

-2835-

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2845

- 5 A DNA sequence (GASx1667R) was identified in *S.pyogenes* <SEQ ID 8189> which encodes the amino acid sequence <SEQ ID 8190>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2967(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 2846

A DNA sequence (GASx1672) was identified in *S.pyogenes* <SEQ ID 8191> which encodes the amino acid sequence <SEQ ID 8192>. Analysis of this protein sequence reveals the following:

Possible site: 24

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -3.82 Transmembrane 3 - 19 (1 - 20)

----- Final Results -----

bacterial membrane --- Certainty=0.2529(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

- 35 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2847

A DNA sequence (GASx1673R) was identified in *S.pyogenes* <SEQ ID 8193> which encodes the amino acid sequence <SEQ ID 8194>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -8.86 Transmembrane 51 - 67 (47 - 75)
INTEGRAL Likelihood = -5.20 Transmembrane 27 - 43 (24 - 45)
INTEGRAL Likelihood = -3.66 Transmembrane 112 - 128 (112 - 131)

----- Final Results -----

-2836-

bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF41294 GB:AE002440 conserved hypothetical protein [Neisseria
 meningitidis MC58]
 Identities = 61/148 (41%), Positives = 96/148 (64%)
 10 Query: 1 LKKSITNEKAILAQGGQEFGAQNTKFLTLHIMIYVFAVIEALLKQIKFDGISFLGLLLM 60
 L SI +EKA++A+G +++G N+ L +H + Y+ + L F+GIS +G L +
 Sbjct: 19 LAVSIKHEKALIAKGAKQYGKTNSTLLAAVHTLYLACFVWVWLSDTAFNGISLIGTLTV 78
 15 Query: 61 LLSVAVLYEVTRILGDIWTVKMLAKDHKYVDHWLFKTIKHPNYFLNIAPELVGIALLC 120
 + S +L + + LG+IWTVK+ + +H+ WLFKT +HPNYFLNI PEL+GIALLC
 Sbjct: 79 MASFVILSLIKQLGEIWTVKIYILPNHQINRSWLFKTFRHPNYFLNIPELIGIALLCQ 138
 Query: 121 AKITAMLLFPCYIVVIYLRIREENKLLA 148
 20 A ++ P Y++V++ RIR+E + +A
 Sbjct: 139 AWYVLLIGLPIYLLVLFKRIRQEEQAMA 166

A related GBS gene <SEQ ID 9009> and protein <SEQ ID 9010> were also identified. Analysis of this protein sequence reveals the following:

25 Lipop: Possible site: -1 Crend: 0
 McG: Discrim Score: 5.86
 GvH: Signal Score (-7.5): 0.14
 Possible site: 60
 >>> Seems to have a cleavable N-term signal seq.
 30 ALOM program count: 2 value: -8.23 threshold: 0.0
 INTEGRAL Likelihood = -8.23 Transmembrane 69 - 85 (64 - 89)
 INTEGRAL Likelihood = -3.29 Transmembrane 142 - 158 (140 - 159)
 PERIPHERAL Likelihood = 1.70 123
 modified ALOM score: 2.15
 35 *** Reasoning Step: 3
 ----- Final Results -----
 40 bacterial membrane --- Certainty=0.4291(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

42.1/64.0% over 168aa
 45 imported
 EGAD|177248| conserved hypothetical protein {Neisseria meningitidis} Insert characterized
 GP|7379797|emb|CAB84365.1|AL162755 putative integral membrane protein {Neisseria
 meningitidis} Insert characterized
 50 GP|7226121|gb|AAF41294.1|AE002440 conserved hypothetical protein {Neisseria meningitidis
 MC58} Insert characterized
 PIR|F81147|F81147 probable integral membrane protein NMA1102 - Neisseria meningitidis
 (group B strain MD58, group A strain Z2491) Insert
 characterized
 55 ORF00432(301 - 807 of 1140)
 EGAD|177248|NMB0883(1 - 169 of 169) conserved hypothetical protein {Neisseria
 meningitidis}GP|7379797|emb|CAB84365.1|AL162755 putative integral membrane protein
 {Neisseria meningitidis}GP|7226121|gb|AAF41294.1|AE002440 conserved hypothetical protein
 {Neisseria meningitidis MC58}PIR|F81147|F81147 probable integral membrane protein NMA1102
 60 [imported] - Neisseria meningitidis (group B strain MD58, group A strain Z2491)
 %Match = 19.0
 %Identity = 42.0 %Similarity = 63.9
 Matches = 71 Mismatches = 61 Conservative Sub.s = 37

-2837-

```

237      267      297      327      357      387      417      447
SSGEYHLLTSDHSLV*IGKAXX*LIXXXEFTMSIIIGLMAAMFIIRLAYLKLSIANEKALRKNGAKEYGVGVSKAITVLH
5      |::|: :: | || | : | : | | | | | : | : |
      MTMILSILSLFFIIRILLFLAVSIKHEKALIAKGAKQYGKTNSTLLAAVH
      10      20      30      40

477      507      537      567      597      627      657      687
IIIFYFSSVTEAILTKASFNFVSVIGLSLMIFSVFMLHTVTRILLGRWTVKLMVDKNHQFVDHHLFRVVKHPNYFLNIAPE
10      : |:: | : : | | : : | : : | | | | : : | | | | : : | | | | | | | |
      TLYYLACFVWVWLSDTAFNGISLIGTLTVMASFVILSLIIKQLGEIWTVKIYILPNHQINRSWLFKTFRHPNYFLNIIPE
      60      70      80      90      100      110      120

717      747      777      807      837      867      897      927
LIGVTLLCHAKYTALFVLPIYAFVIYLRIRENLLLKTIIPNGIKKSRVY*E*DK**T*KSEFFVILSQ*EEVFISCFSS
15      |::| | | | | : : : | | : | :
      LIGIALLCQAWVLLIGLPIYLLVLFRIRIQEEQAMATLF
      140      150      160

```

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2848

A DNA sequence (GASx1674R) was identified in *S.pyogenes* <SEQ ID 8195> which encodes the amino acid sequence <SEQ ID 8196>. Analysis of this protein sequence reveals the following:

```

25      Possible site: 23

      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
30      bacterial cytoplasm --- Certainty=0.3098(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

35 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2849

40 A DNA sequence (GASx1677R) was identified in *S.pyogenes* <SEQ ID 8197> which encodes the amino acid sequence <SEQ ID 8198>. Analysis of this protein sequence reveals the following:

```

Possible site: 33

>>> Seems to have no N-terminal signal sequence

45    INTEGRAL    Likelihood = -8.86    Transmembrane    254 - 270 ( 248 - 280)
      INTEGRAL    Likelihood = -7.01    Transmembrane    303 - 319 ( 296 - 322)
      INTEGRAL    Likelihood = -2.39    Transmembrane    74 - 90 ( 74 - 91)
      INTEGRAL    Likelihood = -1.91    Transmembrane    201 - 217 ( 199 - 217)
      INTEGRAL    Likelihood = -1.91    Transmembrane    223 - 239 ( 220 - 240)
50    INTEGRAL    Likelihood = -1.65    Transmembrane    118 - 134 ( 115 - 135)
      INTEGRAL    Likelihood = -1.49    Transmembrane    56 - 72 ( 55 - 72)
      INTEGRAL    Likelihood = -0.32    Transmembrane    13 - 29 ( 13 - 30)

----- Final Results -----
55    bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-2838-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

5  >GP:BAB05126 GB:AP001511 unknown conserved protein [Bacillus halodurans]
    Identities = 249/534 (46%), Positives = 380/534 (70%)

Query: 12  QDIAFHFFGGLGLFLFSIKYMGDGLQQAAGDKLRYIIDKYTSNPFPGILVGIAMSALIQS 71
          Q + F FFGGLG+FLF IKYMGDGLQ+ AG++LR +DK+T+NP G+L GI ++ L+Q+
10  Sbjct: 6  QTLFLMFFGGLGIFLFGIKYMGDGLQKVAGERLRDLDDKFTTNPLMGVLAVIVTVLLQT 65

Query: 72  SSGVTVTITVGLVSAGLLNLRQAIGIVMGANIGTTTISFLIGFKLGDYALPMIFIGAACLF 131
          S+G TV+T+GLV+AG + L+QAIG++MGANIGTT+T+F+IG K+ +YALP+I +GAA +F
15  Sbjct: 66  STGTTVLTITGLVNAGFMTLKQAIGVIMGANIGTTVTAFTIGIKISEYALPIIIVGAALIF 125

Query: 132  FTSNKKLNNFGRIFGVGGIFFSLNLMGDAMDPLKSVSAFQNYLATLGDKPFQGVFIGTA 191
          F NKK+NN G++IFG G +F+ LN MG+ ++PL+ + AF + ++ + P GV IGT
20  Sbjct: 126  FIKNKKVNNIGQVIFGFGTLFYGLNTMGEGLNPLRELQAFADLTVSMSENPLLGVLTIGTI 185

Query: 192  LTMLIQSSAAIIGILQGLFSGGLLTQGAIPILGSGNIGTCITAVLAAIGSNIAAKRVAA 251
          T +QSS+A IG+LQ L+ G + L A+P+L G NIGT ITAVLAAIG+++AAKR A
25  Sbjct: 186  FTAAVQSSASIGLLQQLYDQCAMDLFAALPVLFGDNIGTTITAVLAAIGASVAAKRAAL 245

Query: 252  AHVLFNLIGTIIIFMIILVPFTSLMLWLQSKLSLTPEMTIAFSHGSFNITNTIILLIPFISL 311
          HV+FNLTIGTII +II++PFT + +L +L MTIAF+HG FN++NTI+ PFI +
30  Sbjct: 246  THVIFNLIGTIIIVLIIIPFTHFIAYLAEVFNLRPMTIAFAHGIFNVSNITIIQFPFIGI 305

Query: 312  LAMIVTRLIPGEDEVVYKYEALYDRLLITQAPSIAGNAHKELVHLASYAIQAFEASYSY 371
          LA+IVT+L+PG+D ++Y+A +LD + +P+IALG A +E++ +A ++ + Y
35  Sbjct: 306  LAIIVTKLVPGDDFYIEYKAKHLDPFRVGSSPAIALGQAKQEVLRMAEFSEKGLLEVSKY 365

Query: 372  IMTADGKFGKVKRYERAVIDTIDEELTTYLVDSNEALSPSENEVLGILDSSRDLERIG 431
          + K E ++E A++ +D ++T YL+ IS+ +LS +++++ ++D+ RD+ERIG
40  Sbjct: 366  MENGQKHAEMAVQFEDAINNLDKITEYLISISSRSLSAQDSKMHGMLMDTVRDIERIG 425

Query: 432  DHSESLGILIEGIISKQIGFSISARQELTEMYQLTHCLTLDAIRAIVDSDTDLAQTIVTR 491
          DH E++ L + + ++ S A +L EM+ LTH +AI ++ D + A++++ +
45  Sbjct: 426  DHINIVELKDYQKANKVKISEKALHDLQEMFDLTHSTLTEAIMSLETGDLEAARSVIEK 485

Query: 492  HKEIEEKERRLRKTHIKRLNCGECTAQAGINFIDIISHYTRITDHALNLAEKVL 545
          + I++ ER+LRK HI R+N G CT AGI F+DI+S+ RI DH++N+AE V+
50  Sbjct: 486  EEHIDQMERKLKQHIIRVNEGNCCTGAAGIVFVDIVSNLERIGDHSVNIAEAVI 539

```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2850

A DNA sequence (GASx1678R) was identified in *S.pyogenes* <SEQ ID 8199> which encodes the amino acid sequence <SEQ ID 8200>. Analysis of this protein sequence reveals the following:

```

Possible site: 48

50  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2940(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
55  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2839-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2851

A DNA sequence (GASx1685R) was identified in *S.pyogenes* <SEQ ID 8201> which encodes the amino acid sequence <SEQ ID 8202>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -7.11 Transmembrane 13 - 29 (9 - 31)

----- Final Results -----

bacterial membrane --- Certainty=0.3845(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2852

A DNA sequence (GASx1695R) was identified in *S.pyogenes* <SEQ ID 8203> which encodes the amino acid sequence <SEQ ID 8204>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1357(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2853

A DNA sequence (GASx1698) was identified in *S.pyogenes* <SEQ ID 8205> which encodes the amino acid sequence <SEQ ID 8206>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1970(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

-2840-

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2854

- 5 A DNA sequence (GASx1713) was identified in *S.pyogenes* <SEQ ID 8207> which encodes the amino acid sequence <SEQ ID 8208>. Analysis of this protein sequence reveals the following:

Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3092(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 2855

A DNA sequence (GASx1737) was identified in *S.pyogenes* <SEQ ID 8209> which encodes the amino acid sequence <SEQ ID 8210>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1878(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2856

A DNA sequence (GASx1748R) was identified in *S.pyogenes* <SEQ ID 8211> which encodes the amino acid sequence <SEQ ID 8212>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2841(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2841-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2857

A DNA sequence (GASx1750R) was identified in *S.pyogenes* <SEQ ID 8213> which encodes the amino acid sequence <SEQ ID 8214>. Analysis of this protein sequence reveals the following:

Possible site: 59

```

10  >>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -1.22      Transmembrane      18 - 34 ( 18 - 34)

      ----- Final Results -----
15          bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

20 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2858

A DNA sequence (GASx1754) was identified in *S.pyogenes* <SEQ ID 8215> which encodes the amino acid sequence <SEQ ID 8216>. Analysis of this protein sequence reveals the following:

```

25  Possible site: 44

      >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
30          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

35 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2859

40 A DNA sequence (GASx1759) was identified in *S.pyogenes* <SEQ ID 8217> which encodes the amino acid sequence <SEQ ID 8218>. Analysis of this protein sequence reveals the following:

Possible site: 36

```

      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
45          bacterial cytoplasm --- Certainty=0.1534(Affirmative) < succ>

```

-2842-

```

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

5 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2860

10 A DNA sequence (GASx1764R) was identified in *S.pyogenes* <SEQ ID 8219> which encodes the amino acid sequence <SEQ ID 8220>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have a cleavable N-term signal seq.

```

15 INTEGRAL    Likelihood = -6.74    Transmembrane    90 - 106 ( 87 - 121)
INTEGRAL    Likelihood = -4.57    Transmembrane    210 - 226 ( 205 - 229)
INTEGRAL    Likelihood = -4.19    Transmembrane    43 - 59 ( 42 - 62)
INTEGRAL    Likelihood = -3.77    Transmembrane    137 - 153 ( 137 - 155)

```

----- Final Results -----

```

20 bacterial membrane --- Certainty=0.3697(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

25 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2861

30 A DNA sequence (GASx1768R) was identified in *S.pyogenes* <SEQ ID 8221> which encodes the amino acid sequence <SEQ ID 8222>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have an uncleavable N-term signal seq

```

35 INTEGRAL    Likelihood =-12.37    Transmembrane    26 - 42 ( 17 - 47)
INTEGRAL    Likelihood = -7.54    Transmembrane    53 - 69 ( 46 - 73)
INTEGRAL    Likelihood = -3.29    Transmembrane    209 - 225 ( 209 - 225)
INTEGRAL    Likelihood = -2.13    Transmembrane    82 - 98 ( 82 - 98)
INTEGRAL    Likelihood = -1.65    Transmembrane    9 - 25 ( 9 - 25)
40 INTEGRAL    Likelihood = -0.85    Transmembrane    117 - 133 ( 117 - 134)

```

----- Final Results -----

```

45 bacterial membrane --- Certainty=0.5946(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAB84959 GB:AE000829 conserved protein [Methanobacterium
thermoautotrophicum]

```

50 Identities = 54/192 (28%), Positives = 90/192 (46%), Gaps = 6/192 (3%)

-2843-

Query: 7 TKLLLLVLANACFFFRVDGFLEFIIVIFLLLLLSALNKKKLA--FKLAVVYLIMIGLSVI 64
 +KL ++V A F D L I + + L++ + A F ++ ++ L++I
 Sbjct: 32 SKLTVVVSATLLSTFISDLTLLIIMGVIFTALIAHSGSLRFAAPFLSFILFWLVSLAI 91

5 Query: 65 PLSIFPSYLDHLLSFVSIAGRLVFPSSLAGLITIKTTTTIYELVHGLRKWRFPPEVWLLTLA 124
 + S H + F+S+ F AGL TT +L LR R P + TL
 Sbjct: 92 MVL---SGNPHTMGFLSLFFARFFIISAAGLSFAFTTEPQKLAESLRSVRIPGEIVFTLT 148

10 Query: 125 VMCRFIPMIRQECCVIHRSKIRGIILTKWSILIRPKQYLEYLMVPLLLSLIRSSQELTI 184
 V R+IP + E I SLK+R L+ SI+ RP L++P+++ ++ S E+ I
 Sbjct: 149 VALRYIPALAVEASSIWDLSKLR-TSLSGSSIIRRPSTLYRGLIIPMIIRTVKISDEVAI 207

Query: 185 ASLTKGLAVNKG 196
 A+ T+G +G
 15 Sbjct: 208 AAETRGFNPREG 219

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2862

20 A DNA sequence (GASx1769R) was identified in *S.pyogenes* <SEQ ID 8223> which encodes the amino acid sequence <SEQ ID 8224>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have a cleavable N-term signal seq.

25 INTEGRAL Likelihood = -7.32 Transmembrane 164 - 180 (158 - 186)
 INTEGRAL Likelihood = -4.67 Transmembrane 85 - 101 (84 - 105)
 INTEGRAL Likelihood = -3.03 Transmembrane 42 - 58 (42 - 61)
 INTEGRAL Likelihood = -2.76 Transmembrane 118 - 134 (117 - 134)
 30 INTEGRAL Likelihood = -2.07 Transmembrane 64 - 80 (64 - 82)
 INTEGRAL Likelihood = -1.22 Transmembrane 18 - 34 (17 - 34)

----- Final Results -----
 bacterial membrane --- Certainty=0.3930(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 40 antigens for vaccines or diagnostics.

Example 2863

A DNA sequence (GASx1776R) was identified in *S.pyogenes* <SEQ ID 8225> which encodes the amino acid sequence <SEQ ID 8226>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have an uncleavable N-term signal seq

45 INTEGRAL Likelihood = -6.37 Transmembrane 4 - 20 (1 - 22)
 INTEGRAL Likelihood = -0.43 Transmembrane 261 - 277 (261 - 278)

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.3548(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 No corresponding DNA sequence was identified in *S.agalactiae*.

-2844-

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2864

- 5 A DNA sequence (GASx1777R) was identified in *S.pyogenes* <SEQ ID 8227> which encodes the amino acid sequence <SEQ ID 8228>. Analysis of this protein sequence reveals the following:

Possible site: 24

```

10 >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -8.17    Transmembrane 1217 -1233 (1215 -1235)

    ----- Final Results -----
    bacterial membrane --- Certainty=0.4270(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

20 >GP:AAF53254 GB:AE003639 CG16974 gene product [Drosophila
    melanogaster]
    Identities = 84/238 (35%), Positives = 133/238 (55%), Gaps = 10/238 (4%)

    Query: 516 LRLDHYELTDISLL--KHAKNITELHLDGNQITEIPKELFSQMKQLRFLNLRSNHLTYLD 573
           L +   L++ SLL ++ K + ELHLD +++T +P+   ++ +LR LNL N LT L
25 Sbjct: 232 LEMSGNRLSNCSLLNLQYMKQLQELHLDRELTYPQRFGLGELSELRLMLNLSQNLLTELP 291

    Query: 574 KDTFKSNAQLRELYLSSNFIHSLEGGFLFQSLHHLEQLDLSKNRIGRLCDNPFEGLSRLTS 633
           +D F   +L L YLS N + L LFQ+ L+ LDLS NR+   DN F   +L
30 Sbjct: 292 RDIFVGALKLERLYLSGNRLSVLPFMLFQTAADLQVLDLSDNRLLSFPDNFFARNGQLRQ 351

    Query: 634 LGFAENSLEEIPEKALEPLTSLNFIIDLSQNNLALLP-KTIEKLRALSTIVASRNHITRID 692
           L   N L+ I + +L L L +DLSON+L+++ K E L L + S N++T +
35 Sbjct: 352 LHLQRNQLKSIGKHSLSYSLRELRLQDLSONSLVIDRKAFESLDHLLALNVSGNNLTLLS 411

    Query: 693 NISFKNLPKLSVLDLSTNEISNLPNGIFKQNNQL-----TKLDDFFNNLLTQVEESV 743
           +I F++L L LDLS N+ LP+G+F++ L   T ++ F+N +++ +ES+
    Sbjct: 412 SIIFQSLHALRQLDLSRNQFKQLPSGLFQRQSLVLLRIDETPIEQFSNWISRYDESL 469

```

- 40 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2865

A DNA sequence (GASx1778R) was identified in *S.pyogenes* <SEQ ID 8229> which encodes the amino acid sequence <SEQ ID 8230>. Analysis of this protein sequence reveals the following:

```

45 Possible site: 39

    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.1067(Affirmative) < succ>
50 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

-2845-

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2866

- 5 A DNA sequence (GASx1779) was identified in *S.pyogenes* <SEQ ID 8231> which encodes the amino acid sequence <SEQ ID 8232>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1885(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 2867

A DNA sequence (GASx1786R) was identified in *S.pyogenes* <SEQ ID 8233> which encodes the amino acid sequence <SEQ ID 8234>. Analysis of this protein sequence reveals the following:

Possible site: 19

25 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0612(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

35 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2868

A DNA sequence (GASx1790) was identified in *S.pyogenes* <SEQ ID 8235> which encodes the amino acid sequence <SEQ ID 8236>. Analysis of this protein sequence reveals the following:

40 Possible site: 13

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

45 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2846-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2869

A DNA sequence (GASx1791R) was identified in *S.pyogenes* <SEQ ID 8237> which encodes the amino acid sequence <SEQ ID 8238>. Analysis of this protein sequence reveals the following:

Possible site: 43

```

10  >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.90    Transmembrane    28 - 44 ( 28 - 44)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.1362(Affirmative) < succ>
15      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9155> which encodes the amino acid sequence <SEQ ID 9156>. Analysis of this protein sequence reveals the following:

```

20  Possible site: 25
      >>> Seems to have a cleavable N-term signal seq.

      ----- Final Results -----
      bacterial outside --- Certainty= 0.300(Affirmative) < succ>
25      bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

30  >GP:AAA24923 GB:L06331 endoglycosidase [Chryseobacterium
      meningosepticum]
      Identities = 105/322 (32%), Positives = 153/322 (46%), Gaps = 53/322 (16%)

      Query: 106 ADKQAQELAKMKIPEKIPMKPLHGSLYGGYFRTWHDKTSDPTEKDKVNSMGELPKEVDLA 165
35      A K ++ ++ I K + GY+RTW D T + SM LP +D+
      Sbjct: 37 AQKSGVTSAVNLSNLIAYKNSDQHSAGYYRTWRDSA---TASGNLPSMRWLPDSLDMV 93

      Query: 166 FIFHDWTKDYSLFWKELATKHVPKLNKQGTRVIRTIPWRFLAGGDNNGIAEDTSKYPNTP 225
      +F D+T + +W L T +VP L+K+GT+VI T+ G NS T+
40  Sbjct: 94 MVFPDYTPPENAYWNTLTKTNVVPYLHKGRTKVIITL-----GDLNSA---TTTGGQDS 143

      Query: 226 EGNKALAKAIVDEYVYKYNLDGLDLDVDEHDSIPKVDKKEDTAGVERSIQVFEEIGKLIGP 285
      G + AK I D++V +YNLDG+D+D+E A + + + + + K GP
45  Sbjct: 144 IGYSSWAKGIYDKWVGEYNLDGIDIDIE-----SSPSGATLTKFVAATKALSKYFGP 195

      Query: 286 KGVDKSRFLFMDSTYMADKNP--LIERGAPYINLLLVQVYGSQGEKGGWEPVSNRPEKTM 343
      K + F+ D+ ++NP + AP N + +Q YG R +
50  Sbjct: 196 KS-GTGKTFVYDT---NQNPNTNFFIQTAPRYNYVFLQAYG-----RSTTNL 237

      Query: 344 EERWQGYSKYIRPEQYMIGFSFYEENAEQGNLWYDINSRKDEDKANGINTDITGTRAERY 403
      Y+ YI +Q++ GFSFYEEN GN W D+ + NG TG RA Y
55  Sbjct: 238 TTVSGLYAPYISMKQFLPGFSFYEENGYPGNYWNDVRYPO-----NG-----TG-RAYDY 286

      Query: 404 ARWQPKTGGVKGGIFSYAIDRD 425
      ARWQP T G KGG+FSYAI+RD
      Sbjct: 287 ARWQPAT-GKKGGVFSYAIERD 307

```

-2847-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2870

A DNA sequence (GASx1803) was identified in *S.pyogenes* <SEQ ID 8239> which encodes the amino acid sequence <SEQ ID 8240>. Analysis of this protein sequence reveals the following:

Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2099(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2871

A DNA sequence (GASx1806R) was identified in *S.pyogenes* <SEQ ID 8241> which encodes the amino acid sequence <SEQ ID 8242>. Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2706(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB16126 GB:Z99124 ribosomal protein S18 [Bacillus subtilis]
Identities = 51/77 (66%), Positives = 63/77 (81%)

Query: 1 MAQQRGGGFKRRKKVDFIAANKIEYVDYKDTELLRSFVSERGKILPRRVGTGSAKNQRKV 60
MA RRGG +R+KV + +N I ++DYKD +LL +FVSERGKILPRRVGT+AK QRK+
Sbjct: 3 MAGGRRGGRAKRRKVCYFTSNGITHIDYKDVDLLKKFVSERGKILPRRVGTGNAKYQRKL 62

Query: 61 TTAIKRARVMALMPYVN 77
T AIKRAR MAL+PYV+
Sbjct: 63 TAAIKRARQMALLPYVS 79

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2872

A DNA sequence (GASx1809R) was identified in *S.pyogenes* <SEQ ID 8243> which encodes the amino acid sequence <SEQ ID 8244>. Analysis of this protein sequence reveals the following:

Possible site: 60

-2848-

```

>>> Seems to have an uncleavable N-term signal seq
  INTEGRAL    Likelihood = -7.59    Transmembrane  70 - 86 ( 66 - 92)
  INTEGRAL    Likelihood = -6.42    Transmembrane  13 - 29 (  8 - 33)
  INTEGRAL    Likelihood = -5.68    Transmembrane  48 - 64 ( 43 - 69)

----- Final Results -----
      bacterial membrane --- Certainty=0.4036(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2873

A DNA sequence (GASx1813R) was identified in *S.pyogenes* <SEQ ID 8245> which encodes the amino acid sequence <SEQ ID 8246>. Analysis of this protein sequence reveals the following:

```

Possible site: 56

>>> Seems to have a cleavable N-term signal seq.
  INTEGRAL    Likelihood =-10.51    Transmembrane  127 - 143 ( 113 - 147)
  INTEGRAL    Likelihood =-10.46    Transmembrane  151 - 167 ( 149 - 167)
  INTEGRAL    Likelihood = -4.41    Transmembrane   59 - 75 (  57 - 77)

----- Final Results -----
      bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAB98363 GB:U67490 lipoprotein B (lppB) [Methanococcus
  jannaschii]
Identities = 43/143 (30%), Positives = 68/143 (47%), Gaps = 7/143 (4%)

Query: 25  LLNVLLKIITGVMY--ILYPSFLIFTLWQGMTFQLWLRLLIIPAVGFIALSYIRKRFDFP 82
      + + ++ II+  Y I  S +IF  +  +L  L +  + F +L Y+  P
Sbjct: 181  IFDAIMPIISKATAYPLIAITSLIIFIKNRKFGMKLIFALFLAFMIAF-SLKYLVE---P 236

Query: 83  RPYEKWNIKPLIDKDTKGRSMPSRHVFSATMISMCLLRYVYVFGIVCLILSALLAICRVI 142
      RPY  +  L+  +  S PS H  A ++ LL Y  GI+ L  + ++A RV
Sbjct: 237  RPYLVLDNVHLLCNEGNEPSFSPSGHTTLAFTLATSLLFYSKKLGLFLSWAIIIVAYSRYV 296

Query: 143 AGIHYPKDVIVGYLIGLMLGLCL 165
      G+HYP DV+ G +IG+  G CL
Sbjct: 297 VGVHYPLDVLAGMIIGIFCG-CL 318

```

A related GBS gene <SEQ ID 9011> and protein <SEQ ID 9012> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1  Crend: 9
McG: Discrim Score:      3.19
GvH: Signal Score (-7.5): -2.18
      Possible site: 55
>>> Seems to have a cleavable N-term signal seq.
ALOM program  count: 3 value: -11.78 threshold:  0.0

```

-2849-

```

INTEGRAL    Likelihood =-11.78    Transmembrane  126 - 142 ( 112 - 147)
INTEGRAL    Likelihood =-11.30    Transmembrane  150 - 166 ( 147 - 166)
INTEGRAL    Likelihood = -4.41    Transmembrane   58 -  74 (  56 -  76)
PERIPHERAL  Likelihood =  3.29      107
5 modified ALOM score:  2.86

```

*** Reasoning Step: 3

----- Final Results -----

```

10      bacterial membrane --- Certainty=0.5713(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

15 ORF01020(472 - 792 of 1098)
EGAD|44548|MJ0374(213 - 318 of 330) conserved hypothetical protein {Methanococcus
jannaschii} OMNI|MJ0374 conserved hypothetical protein SP|Q57819|Y374_METJA HYPOTHETICAL
PROTEIN MJ0374. GP|1591081|gb|AAB98363.1||U67490 lipoprotein B (lppB) {Methanococcus
20 jannaschii} PIR|F64346|F64346 hypothetical protein MJ0374 - Methanococcus jannaschii
%Match = 6.8
%Identity = 30.8 %Similarity = 53.3
Matches = 33 Mismatches = 49 Conservative Sub.s = 24

222      252      282      312      342      372      402      432
25 EGVTKYLRRNKHVKHFAYAPQNAGGSGATIVTLG*IMESYEQFYAKLSQPFRKSPQLIILLNFLKIVTGMMYILYPSFL

VIAWLSGIFEMHKLLFTVGTIIGRLPRFLAVAYFGDVLGNINRLSDINIYLFYLINSHYNYIFDAIMPIISKTAIPLIAI
      130      140      150      160      170      180      190

30 462      492      522      552      582      612      642      672
IFTLWQGMTFQLWLRLLIIPAVGFIALSYIRKRLDFPRPYEKWNIKPLIYKDTGRSMPSRHVFSATMISMCLLRYVYF
      ::|:  : |:  :: :: |||  :  |:  :  |||  |  :: |||  :
TSLIIFIKNRKFGMKLIIFALFLAFMIAFSLKYLNEPRPYLVLDNVHLLCNEGNEPSFPSPGHTTLAFTLATSLLFYSKKL
      210      220      230      240      250      260      270

35 702      732      762      792      822      852      882      912
GIVCLILSVLLAICRVIAGIHYPKDVIVGYLIGLILGLCLFI*RVRSK*FQKQLDSCTIGLSLR*NGEKRW*H*K*QMLHL
||: |  ::::|  ||  |:|||  ||: |  :::|  |||
40 GILFLSWAIIIVAYSRYVYGVHYPLDVLAGMIIGIFCG-CLTRIDIYKLIDNI
      290      300      310      320      330

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2874

45 A DNA sequence (GASx1815R) was identified in *S.pyogenes* <SEQ ID 8247> which encodes the amino acid sequence <SEQ ID 8248>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

50      bacterial cytoplasm --- Certainty=0.0888(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
55

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2850-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2875

A DNA sequence (GASx1825R) was identified in *S.pyogenes* <SEQ ID 8249> which encodes the amino acid sequence <SEQ ID 8250>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.16 Transmembrane 7 - 23 (7 - 23)

----- Final Results -----

bacterial membrane --- Certainty=0.1065(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2876

A DNA sequence (GASx1832) was identified in *S.pyogenes* <SEQ ID 8251> which encodes the amino acid sequence <SEQ ID 8252>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0918(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2877

A DNA sequence (GASx1836R) was identified in *S.pyogenes* <SEQ ID 8253> which encodes the amino acid sequence <SEQ ID 8254>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4084(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

-2851-

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2878

- 5 A DNA sequence (GASx1864R) was identified in *S.pyogenes* <SEQ ID 8255> which encodes the amino acid sequence <SEQ ID 8256>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5280(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC36810 GB:L12244 ribosomal protein L28 [Bacillus subtilis]
Identities = 45/62 (72%), Positives = 52/62 (83%)

Query: 1 MAKVCYFTGRKTVSGNNRSHAMNQTKRTVKPNLQKVITILVDGKPKKVWASARALKSGKVE 60
MA+ C TG+KT +GNNRSHAMN +KRT NLQKV ILV+GKPKKV+ SARALKSGKVE
Sbjct: 1 MARKCVITGKKT TAGNNRSHAMNASKRTWGANLQKVRILVNGKPKKVYVSARALKSGKVE 60

Query: 61 RI 62
R+
Sbjct: 61 RV 62

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2879

A DNA sequence (GASx1869) was identified in *S.pyogenes* <SEQ ID 8257> which encodes the amino acid sequence <SEQ ID 8258>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1858(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

- 45 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2852-

Example 2880

A DNA sequence (GASx1881) was identified in *S.pyogenes* <SEQ ID 8259> which encodes the amino acid sequence <SEQ ID 8260>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2752(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif 136-138

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF04356 GB:AF177167 type IC restriction subunit [Streptococcus thermophilus]
Identities = 358/1047 (34%), Positives = 571/1047 (54%), Gaps = 91/1047 (8%)

Query: 7 TELELEKELIHLLLETGESQWYRKELKTEDALWDFKILAQNNQYLNNEPLTASEKEQ 66
+E +E + I +L E+QWYR +LK+E+ALW NF L + N L E+PLT E +Q
Sbjct: 4 SEQMIENQFIQILSEKENQWYRDLKSEELWQNFRLNRLNLAFLGEQPLTDKEFKQ 63

Query: 67 IKNQNLNFVNY--YEAACKWLAGENGIKQVQREDAKLGTIRLEVVKADNVAGGTSVYEIA 124
+K + + + + A++WL GENG+A++ ++RED K + LE + +++GGTS YE+
Sbjct: 64 VKVEFSRLTGTPFLASQWLRGENGVQAILLEREDGK--RVTLEAFRNKDISGGTSSYEVV 121

Query: 125 NQVAFSGSRDRRGDVTLLINGLPMIQIELKSNHQ--CIEAFNQVKYDKEGQFRGIFST 182
+QV SR RGDV+LLINGLP+I IELK ++ + ++A+ Q+++Y ++G F+GI++T
Sbjct: 122 HQVVPDSSRVDRGDVSLINGLPIIHIELKKESAKDGFMAQYQIQRYAEDGDFKGIYAT 181

Query: 183 LQMFVVSNKTDTRYIAAANKLNLP-----NFLTQWVDQNNKPKDLFAFAKEVLSIPRA 237
Q+ V+SNK DTRY A E+ FL W ++N+ DLF F + VL IP A
Sbjct: 182 TQIMVISNKVDTRYFARPS EDTAEAYARMKKFLFNWRTEDNQTVSDLDFTRTVLRIPDA 241

Query: 238 HQMVMTYSVIDDDKKA---LILLRPYQIHAI EAVAEASRHRKSGYIWHITGSGKTLTSYK 294
H+++ Y+++ DD+K L+ LRPYQIHAI + + + + G+IWH TSGSKT+TS+
Sbjct: 242 HELISQYTLVDDQKNQKFLMALRPYQIHAIKIRKQAAQHEGGFIWHATGSGKTITSFV 301

Query: 295 VARNILQIP-AVEKSIFVIDRKDLNDQTASAFQSYA-----QNDIFD--VDETEDT 342
+ + Q V++++ V+DR DLD QT F +A +N + + + ++
Sbjct: 302 ATKLLAQNAIGVDRTVMVVDRTDLDAQTQDEFTKFASEYHTGQTENSANTLIVGIKNQ 361

Query: 343 RQLIKNLESS--DRRVVVTTIQKLNAMISQMESYDTPKFKKLKERLAHLNVVVFVDECHR 400
+QL +NL SS + ++VTTIQKL+A + + K E+L ++VF+VDE HR
Sbjct: 362 KQLAQNLSSKNNTILVTTIQKLSAAMRSAQGESEKGSNQFEKLRQEHIVFIVDEAHR 421

Query: 401 AVTPERQRYLTNTFRNSRWYGFGTPIFVENKRAQLGDLAQTTEQQYQKCLHQYTVKEAI 460
AV+ E + + NS W+G TGTPIF ENK+ + G A+TT QQYG LH YT+K A+
Sbjct: 422 AVSDEEMKRIKILPNSTWFGLTGTPIFEENKKQENGTFARTTSQQYGPLLHSTYIKNAM 481

Query: 461 HDKAVLGFQVEYKTTIPD-----MPEDS-----IPEEAYDHEEHMLAVLD 500
D AVLGFQVEY + I + +P+D+ +P E Y+ +EH+ +L
Sbjct: 482 DDGAVLGFQVEYHSLISEEDQEVIVTQLNKGKLPDDALQQEKLPLTELYETDEHIRTMLQ 541

Query: 501 SIINQSR--KKLGFNNGIGQTFEGLLTVKSIARAQAYYDLMKKVKAGETDLVISKVKKEK 558
I N+ KK NG T +LT SIA+A+ Y ++K++K T L+ ++ E+
Sbjct: 542 KIFNRRSVVKFKVKNGF-PTMSAILTTHSIAQAKHIYRILKEMKDNGT-LLNGRQFDER 599

Query: 559 L----PDFPKVAITYSITENDNASISRQDKMTKNLEDYNHLFGTNFTIDNLQGYNRDLND 614
DFP+VAIT+S + + D++ + +++Y F + D + YN+++N
Sbjct: 600 HQLIDKDFPRVAITFTSTNPDLQEKNEQDELVEIMKEYEKQFDASPYQDE-KLYNQININK 658

Query: 615 RLARKKDKFKDRHEQLDLVIVVDRLLTGFDAPCLSTIFIDRQPMKPQHIIQAFSRTNRI 674

-2853-

5
 10
 15
 20
 25

```

    RLARK+ +++ + LD VIVVDRLLTGFD+P + T++IDR+ M Q ++QAFSRTNRI+
Sbjct: 659 RLARKEKQYQSDGQWLD FVIVVDRLLTGFDSP TIQTLYIDRE-MNYQKLLQAFSRTNRIY 717

    Query: 675 ESRKHYGQVVTFQTPLRFKEAVDKALSLYSNGGEN-DVLAP-SWEEEEKARFFEKVTVLKN 732
    + K G +V+F+ P +E V L+SN +N D L P +EE K F E T+ K
Sbjct: 718 -TGKDSGLIVSFRKPFTMRENVNRTFR LFSNEKQNFQDLIPKEYEEVKKEFIECSTLYKQ 776

    Query: 733 IVPDPDAFPTTIESAQTAFLKQYAKAFQAFDKLFASVQVYSDFNETLLSEVGLSDEVIDTY 792
    D P A + Y K +++ L + Q DF E SEV E + Y
Sbjct: 777 SEADLSDNPNDLKTMIAQVSAYQKLEKSYKALRSYDQYEEDEFEE--PSEV---VEQLPQY 831

    Query: 793 KGTYQNVAIEIRKRRED-----DEAIPEINIDYELESVQMDINHYIILTLIQAFVD 844
    +G +N+ +I++ ED ++ + EI +L + D ++ YI L++A
Sbjct: 832 QGKTENIKTKIKEMIEDEGHPEEDFEKLLQEI AFSSQLNATHKDVVDSFYINQLLKAIQL 891

    Query: 845 QEQEALQERLNDNPMDQYIQDLAKSNPAMADSLAELWQDIQKEPKAYEGKSIVYELDNLI 904
    E A+++ + + Q + K + D L ++I + + I
Sbjct: 892 NEAGAVEK--FEKEIQKQDPQIQKMYHTLKDQLVNTTEEI-----DVAQLKETSI 939

    Query: 905 GDKIQRAIKHFADQWKADPKLAFVATNYHSANSTKQVGMSTLKE-SLDYQAYKEKQGDS 963
    ++IQR ++ A+++ D L Y S T L +L + ++ K G+
Sbjct: 940 QNEIQRLQKEAEFGLSFDLQSAMNEYQSDKKTIPYLTHLLDSMTLSKEEF EAKTGE- 998

    Query: 964 AMNKLKYKSQFERELVQFIRDQIQPLK 990
    K + +++ E +Q +Q+Q K
Sbjct: 999 ---KYRRRTKVLEERLQQNF EQLKWK 1022
  
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 2881

A DNA sequence (GASx1882) was identified in *S.pyogenes* <SEQ ID 8261> which encodes the amino acid sequence <SEQ ID 8262>. Analysis of this protein sequence reveals the following:

35
 40

```

Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
    bacterial cytoplasm --- Certainty=0.3653(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
  
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

45
 50
 55
 60

```

>GP:AAB53491 GB:U35629 unknown [Lactococcus lactis subsp. lactis]
Identities = 141/241 (58%), Positives = 178/241 (73%)

    Query: 3 KSKQPQYRFDGFEGEWEEKELGDIVQITMGQSPSSQNYTINPSDYILVQGNADIKNGYVF 62
    K K P+ RF GF EWE ++LGD V+I MGQSP+S+NYT +P+DYILVQGNAD+KNG V
Sbjct: 13 KKKVPELRFKGFTEWELRKLGDVIRIVMGQSPNSENYTDPPNDYILVQGNADMKNRVL 72

    Query: 63 PRVWTTQITKQADKGDIIILSVRAPVGDVGKTNHYHVIIGRGVAAIKGNEFIFQILKYLKEI 122
    PRVWTTQ+TKQA+K D+ILSVRAPVGD+GKT Y V+IGRGVAAIKGNEFIFQ L +K
Sbjct: 73 PRVWTTQVTKQAEKDDLIILSVRAPVGDIGKTAYDVVIGRGVAAIKGNEFIFQNLGKMKSD 132

    Query: 123 GYWKRISTGSTFDSISSSDIKYAKIQIPSLPEQEAIGELFQMVDQLIQLQDQKLATLKEQ 182
    GYW R STGSTF+SI+S+DIK A I +P++ EQ+ IG F+ +D I L +KL LKEQ
Sbjct: 133 GYWTRYSTGSTFESINSTDIKEAIIISVPAIEEQDKIGSFQKQLDNTIALHQRKLDLLKEQ 192

    Query: 183 KQTFRLKMFPAQGGQKVPEIRLQGFKEGEWEEKLREVSTHRSQTAEKYFDSEGEFKVISIG 243
    K+ FL+KMFP G KVPE+R GF +WEE+KL +++ +G G++ + G
Sbjct: 193 KKGFLQKMF PKNGAKVPELRFAGFADDWEERKLG DITKISTGKLDANAMVENGKYDFYTSG 253
  
```

-2854-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2882

- 5 A DNA sequence (GASx1883) was identified in *S.pyogenes* <SEQ ID 8263> which encodes the amino acid sequence <SEQ ID 8264>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4318(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF04357 GB:AF177167 type IC modification subunit [Streptococcus thermophilus]
Identities = 293/523 (56%), Positives = 377/523 (72%), Gaps = 6/523 (1%)

Query: 6 TSLRQALWHSADQLRGQMDANDYKNYLLGLIFYKHLSDKLLAVCDNLEKHFNTFTEAQK 65
TSL Q LW SAD LRG+MDA++YKNYLLGLIFYK+LSDK L V + +TF E
Sbjct: 3 TSLNQQWASADILRGKMDASEYKNYLLGLIFYKYLSDKQLREVYEQENGKTDTFPERST 62

Query: 66 I---FEDAYQDEGLKDDLLISVVTGDLGYFIEPTLTFEKLIQDVYHNTFQLES LAQGFRDI 122
+ F + Y+++ KDDLI + GYFI+P F + F L L GF ++
Sbjct: 63 LYAGFMEWYEED--KDDLIENIQPRQGYFTQPDRLFYHYRIKADNYEFNLTDLQAGFNEL 120

Query: 123 EQSGEDFENLFEDIDLYSKKLGSTPQKQNTISNVMKTLNEIDFEAVDGDTLGDAYEYLI 182
E+ GE+F LF DIDL S KLGS Q++N TI+ V++ L+EID +GD +GDAYEYLI
Sbjct: 121 ERQGEFSGFLFSDIDLNSTKLGSNAQQRNVITITEVLRALDEIDLFEHNGDVIGDAYEYLI 180

Query: 183 GEFASESGKKAGEFYTPQAVSHLMTQIVFLGREDQKGMTLYDPAMGSGSLLLNAKKYSNQ 242
G FA+ +GKKAGEFYTPQAVS +M++I +G+E + +YDPAMGSGSL+LN ++Y
Sbjct: 181 GMFAAGAGKKAGEFYTPQAVSRIMSEITSIGQESRVPFHIYDPAMGSGSLMLNIRRYLIH 240

Query: 243 SDTVSYQQEINTSTYNLARMNMMLHGVAIENQHLSNADTLADWPTEPINFQDGLVLMNP 302
+ V Y+QE+NT+T+NLARMN++LHGV E +L+N DTLADWP++EP FD V+MNP
Sbjct: 241 PNQVHYHGQELNTTTTFNLARMNLILHGVDKERMNLNNGDTLADWPSEEPYQFDSVVMNP 300

Query: 303 PYSLKWSATAGFLTDPRFSSYGVLPKSKADFAFLHGFYHLKNTGTMAIVLPHGVLFVRG 362
PYS KWSA FL+DPRF +G LAPKSKADFAFLHGFYHLK +GTM IVLPHGVLFVRG
Sbjct: 301 PYSAKWSAADKFLSDPRFERFGKLAPKSKADFAFLHGFYHLKESGTMGIVLPHGVLFVRG 360

Query: 363 AAEGKIRQKLEQGAIDTIIGLPNSIFYNTSIPTTTIILKKNRNTKDVFFIDASKFEDKG 422
AEG IRQ LLE GAID +IGLP+NIF+ TSIPTT+IILKKNR+ +DV FIDAS++F+K
Sbjct: 361 GAEGTIRQALLEMGIDAVIDGLPANIFFGTSTIPTTVIILKKNRSDVLFIDASQDFEKQ 420

Query: 423 KNQNTMTDNHIKKILDAYKSRDNDKFSYLSFDEIENDYNLNIPRYVDTFEEVPVKPL 482
KNQN + D HI KI+ YK R++ ++++++ASFDEI END+NLNIPRYVDTFEE L
Sbjct: 421 KNQNVLLDEHIDKIVSTYKKREDIERYAHVASFDEIQENDFNLNIPRYVDTFEEEPVDL 480

Query: 483 PELAKQLSDIDQEIATNAKLDQLMKQLVGTITKEAQDELDTFR 525
E+ L I++E+ + L L+ ++E Q +++ R
Sbjct: 481 VEVNTNLLKINEELVQEQTLTSLINDF-SESEENQAMIESMR 522

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2855-

Example 2883

A DNA sequence (GASx1886R) was identified in *S.pyogenes* <SEQ ID 8265> which encodes the amino acid sequence <SEQ ID 8266>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -8.17	Transmembrane	155 - 171 (147 - 173)
INTEGRAL	Likelihood = -7.22	Transmembrane	14 - 30 (11 - 33)
INTEGRAL	Likelihood = -7.17	Transmembrane	182 - 198 (179 - 205)
INTEGRAL	Likelihood = -5.68	Transmembrane	132 - 148 (128 - 152)
INTEGRAL	Likelihood = -4.14	Transmembrane	46 - 62 (43 - 62)
INTEGRAL	Likelihood = -3.50	Transmembrane	73 - 89 (73 - 90)
INTEGRAL	Likelihood = -0.96	Transmembrane	95 - 111 (95 - 111)

----- Final Results -----

bacterial membrane	---	Certainty=0.4270(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2884

A DNA sequence (GASx1890R) was identified in *S.pyogenes* <SEQ ID 8267> which encodes the amino acid sequence <SEQ ID 8268>. Analysis of this protein sequence reveals the following:

Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.4757(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

RGD motif 339-341

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA62650 GB:L37110 clyM [Plasmid pAD1]

Identities = 127/492 (25%), Positives = 230/492 (45%), Gaps = 30/492 (6%)

Query: 46	KLFYSEFENQLFETIMFLSMKTLVLDINHFSKEIENK----	SEAYEQYIQQ-IREENGIN	100
	K F L + ++ L+ KTLVLD++ F K K S+ + Y+++ + I		
Sbjct: 135	KEFIINLLENLTQELIHLTSKTLVLDLHTFKKNEPLKGNDSKRFIYYLKKRFNSKKDII		194

Query: 101	HFFDRYPYLLKQINKEVGLIEESYSLLFDRFLEDLSEIKSCFNI-	SEPLSNVAFSLGDSH	159
	F+ YP L++ + ++ + R EDL I++CFNI S L++++ S GDSH		
Sbjct: 195	AFYTCYPELMRITVVRMYFLDNTKQMLIRVTEDLPSIQNCFNIQSSELNSISSESQDSH		254

Query: 160	SKKQTVVKIAFKE-KSVYYKPKSYHSHSILLELTSLKSSNIPSFSLPKSLVKADYCWQL		218
	S+ +TV + F + K + YKPK +S + L + L + K + + Y ++		
Sbjct: 255	SRGKTVSTLTFSDGKKIVYKPK- INSENKLRDFFFEFLNKELEADIIYIVKKVTRNTYFYEE		313

Query: 219	GVAYTSSNK-DEVAKIYFKYGVLAASFSEIFSITDLHMENVIVSGGDLYLIDVETFFQRL		277
	+ N +EV K Y +YG L + +F++TDLH EN+I G +ID ETFFQ+ +		
Sbjct: 314	YIDNIEINNIEEVKKYERYGKLGIAFLFNVTDLHYENIIAHGEYPVITIDNETFFQQNI		373

-2856-

Query: 278 NVQONQFEGITVDITYQRIYETSLNGLFP---VQFEKNSAPNVSGISRKGGKRQKGKYEL 334
 ++ N TVD + ++ + GL P ++ + +S +S K Q +++
 Sbjct: 374 PIEFGN--SATVDAKYKYLDSIMVTGLVPYLAMDKSDSKDEGVNLSALNFKEQSVPFKI 431

Query: 335 I---NKNRGDLKLVKVDYFQEDRFNIPTLNGKVVEPLDYANEIISGFRECYIFLLSQRSK 391
 + N +++ + + N P +N + + + Y I++G + + + K
 Sbjct: 432 LKIKNTFTDEMRFEYQTHIMDTAKNTPIMNEKISFISYEKYIVTGMKSILMKAKDSKKK 491

Query: 392 IKEIV-EGFPELKSRVPFRNTSDYGKFLQASTNPKYLFS---EKKRKNLFSILYETKHI 446
 I + L R R T Y L+ S +P + EK N+++ Y+ K +
 Sbjct: 492 ILAYINNNLQNLIVRNVIRPTQRYADMLEFSYHPNCFNSAIEREKVLHNMWAYPYKNKKV 551

Query: 447 EHFIVDNEIKDLMNIGDIP-YFSMDTRGNVNSVGTLLIGNLGDTTSL---FDSITILINDER 502
 H+ E DL++GDIP +++ ++ ++ S G L+ + ++L + I L DE
 Sbjct: 552 VHY---EFSDLIDGDIPIFYNNISKTSLIASDGCLVEDFYQESALNRCINLKNINDLCDED 607

Query: 503 LKFTCELLEIVL 514
 + LEI L
 Sbjct: 608 ISIQTVWLEIAL 619

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2885

A DNA sequence (GASx1891R) was identified in *S.pyogenes* <SEQ ID 8269> which encodes the amino acid sequence <SEQ ID 8270>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3487(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA02867 GB:L07740 salivarin A [Streptococcus salivarius]

Identities = 46/51 (90%), Positives = 48/51 (93%)

Query: 1 MSFMKNSKDILTNAIEEVSEKELMEVAGGKKGSGWFATITDDCPNSVFVCC 51

M+ MKNSKDIL NAIEEVSEKELMEVAGGK+GSGW ATITDDCPNSVFVCC

Sbjct: 1 MNAMKNSKDILNNAIEEVSEKELMEVAGGKRGSGWIATITDDCPNSVFVCC 51

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2886

A DNA sequence (GASx1901R) was identified in *S.pyogenes* <SEQ ID 8271> which encodes the amino acid sequence <SEQ ID 8272>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.59 Transmembrane 3 - 19 (1 - 20)

----- Final Results -----

bacterial membrane --- Certainty=0.1638(Affirmative) < succ>

-2857-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

- 5 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2887

- 10 A DNA sequence (GASx1905R) was identified in *S.pyogenes* <SEQ ID 8273> which encodes the amino acid sequence <SEQ ID 8274>. Analysis of this protein sequence reveals the following:

Possible site: 25

>>> Seems to have an uncleavable N-term signal seq

15 INTEGRAL Likelihood = -0.48 Transmembrane 38 - 54 (37 - 54)

----- Final Results -----

20 bacterial membrane --- Certainty=0.1192 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 2888

A DNA sequence (GASx1911R) was identified in *S.pyogenes* <SEQ ID 8275> which encodes the amino acid sequence <SEQ ID 8276>. Analysis of this protein sequence reveals the following:

Possible site: 30

30 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -10.40 Transmembrane 27 - 43 (22 - 48)

INTEGRAL Likelihood = -9.82 Transmembrane 52 - 68 (50 - 74)

INTEGRAL Likelihood = -7.27 Transmembrane 113 - 129 (111 - 134)

35 INTEGRAL Likelihood = -1.97 Transmembrane 137 - 153 (135 - 153)

----- Final Results -----

40 bacterial membrane --- Certainty=0.5161 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2889

A DNA sequence (GASx1915R) was identified in *S.pyogenes* <SEQ ID 8277> which encodes the amino acid sequence <SEQ ID 8278>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -10.77 Transmembrane 242 - 258 (238 - 262)

----- Final Results -----

bacterial membrane --- Certainty=0.5310(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2890

A DNA sequence (GASx1918R) was identified in *S.pyogenes* <SEQ ID 8279> which encodes the amino acid sequence <SEQ ID 8280>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -7.32 Transmembrane 40 - 56 (39 - 60)

----- Final Results -----

bacterial membrane --- Certainty=0.3930(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2891

A DNA sequence (GASx1923R) was identified in *S.pyogenes* <SEQ ID 8281> which encodes the amino acid sequence <SEQ ID 8282>. Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -12.26 Transmembrane 20 - 36 (13 - 42)

----- Final Results -----

bacterial membrane --- Certainty=0.5904(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2859-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2892

A DNA sequence (GASx1926) was identified in *S.pyogenes* <SEQ ID 8283> which encodes the amino acid sequence <SEQ ID 8284>. Analysis of this protein sequence reveals the following:

Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2322(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2893

A DNA sequence (GASx1928R) was identified in *S.pyogenes* <SEQ ID 8285> which encodes the amino acid sequence <SEQ ID 8286>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3395(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2894

A DNA sequence (GASx1929R) was identified in *S.pyogenes* <SEQ ID 8287> which encodes the amino acid sequence <SEQ ID 8288>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -1.86 Transmembrane 17 - 33 (15 - 33)

----- Final Results -----

bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

-2860-

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2895

- 5 A DNA sequence (GASx1931R) was identified in *S.pyogenes* <SEQ ID 8289> which encodes the amino acid sequence <SEQ ID 8290>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0551(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 2896

A DNA sequence (GASx1941R) was identified in *S.pyogenes* <SEQ ID 8291> which encodes the amino acid sequence <SEQ ID 8292>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2377(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2897

A DNA sequence (GASx1949) was identified in *S.pyogenes* <SEQ ID 8293> which encodes the amino acid sequence <SEQ ID 8294>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0262(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2861-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2898

A DNA sequence (GASx1951R) was identified in *S.pyogenes* <SEQ ID 8295> which encodes the amino acid sequence <SEQ ID 8296>. Analysis of this protein sequence reveals the following:

Possible site: 45

10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.1330 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

20 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2899

A DNA sequence (GASx1953) was identified in *S.pyogenes* <SEQ ID 8297> which encodes the amino acid sequence <SEQ ID 8298>. Analysis of this protein sequence reveals the following:

Possible site: 15

25 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

30 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

35 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2900

A DNA sequence (GASx1957) was identified in *S.pyogenes* <SEQ ID 8299> which encodes the amino acid sequence <SEQ ID 8300>. Analysis of this protein sequence reveals the following:

Possible site: 26

40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2409 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-2862-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2901

A DNA sequence (GASx1969) was identified in *S.pyogenes* <SEQ ID 8301> which encodes the amino acid sequence <SEQ ID 8302>. Analysis of this protein sequence reveals the following:

Possible site: 14

```

10  >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -2.28    Transmembrane    7 - 23 ( 7 - 23)

      ----- Final Results -----
15          bacterial membrane --- Certainty=0.1914(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2902

A DNA sequence (GASx1971R) was identified in *S.pyogenes* <SEQ ID 8303> which encodes the amino acid sequence <SEQ ID 8304>. Analysis of this protein sequence reveals the following:

Possible site: 21

```

25  >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
30          bacterial cytoplasm --- Certainty=0.1545(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2903

A DNA sequence (GASx1973) was identified in *S.pyogenes* <SEQ ID 8305> which encodes the amino acid sequence <SEQ ID 8306>. Analysis of this protein sequence reveals the following:

Possible site: 49

```

40  >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -2.44    Transmembrane    31 - 47 ( 31 - 48)

      ----- Final Results -----
45          bacterial membrane --- Certainty=0.1977(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-2863-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

5  >GP:CAB51744 GB:AJ245405 speX [Streptococcus pyogenes]
    Identities = 236/256 (92%), Positives = 243/256 (94%)

    Query: 3  MIISFESVILKHNKIITPEKRLFMKKTCLIFSFTSIFIAIISRPVFGLEVDNNSLLRNIY 62
              MIISFESVILKHNKIITPEKRLFMKKTCLIFSFTSIFIAIISRPVFGLEVDNNSLLRNIY
10  Sbjct: 1  MIISFESVILKHNKIITPEKRLFMKKTCLIFSFTSIFIAIISRPVFGLEVDNNSLLRNIY 60

    Query: 63  STIVYEYSDTVIDFKTSHNLVTKKLDVRDARDFINSEMDEYAANDFKDGDKIAMFSVPF 122
              STIVYEYSD VIDFKTSHNLVTKKLDVRDARDFINSEMDEYAANDFK GDKIA+FSVPF
15  Sbjct: 61  STIVYEYSDIVIDFKTSHNLVTKKLDVRDARDFINSEMDEYAANDFKTGDKIAVFSVPF 120

    Query: 123 DWNYLSEGVIAITYGGMTPTYQEEPMKSNIPVNLWINRKQIPVPYNQISTNKTTVTAQEI 182
              DWNYLS+GKV AITYGG+TPYQ+ K VNLWIN KQI VPYN+ISTNKTTVTAQEI
20  Sbjct: 121 DWNYLSKGKVTAYTYGGITPYQKLQYLKISLVNLWINGKQISVPYNEISTNKTTVTAQEI 180

    Query: 183 DLKVRKFLISQHQLYSSGSSYKSGKLVFHTNDNSDKYSLDLFYVGYRDKESIFKVYKDNK 242
              DLKVRKFLI+QHLYSSGSSYKSG+LVFHTNDNSDKYS DLFYVGYRDKESIFKVYKDNK
25  Sbjct: 181 DLKVRKFLIAQHLYSSGSSYKSGRLVFHTNDNSDKYSFDLFYVGYRDKESIFKVYKDNK 240

    Query: 243 SFNIDKIGHLDIEIDS 258
              SFNIDKIGHLDIEIDS
30  Sbjct: 241 SFNIDKIGHLDIEIDS 256

```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 2904

A DNA sequence (GASx1974R) was identified in *S.pyogenes* <SEQ ID 8307> which encodes the amino acid sequence <SEQ ID 8308>. Analysis of this protein sequence reveals the following:

```

    Possible site: 53

35  >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2022(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
40  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2905

A DNA sequence (GASx1983) was identified in *S.pyogenes* <SEQ ID 8309> which encodes the amino acid sequence <SEQ ID 8310>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 14

    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----

```

-2864-

```

bacterial cytoplasm --- Certainty=0.0989(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

5 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2906

10 A DNA sequence (GASx1987) was identified in *S.pyogenes* <SEQ ID 8311> which encodes the amino acid sequence <SEQ ID 8312>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.2389(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

20 No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2907

25 A DNA sequence (GASx1988) was identified in *S.pyogenes* <SEQ ID 8313> which encodes the amino acid sequence <SEQ ID 8314>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.5904(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

35 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:BA16031 GB:AB030747 transposase [Streptococcus pyogenes]
Identities = 22/24 (91%), Positives = 23/24 (95%)

```

Query: 1 LERLFGTAKYHNLCTREKGSK 24

+ERLFGTAKYHNL YTREKGSK

Sbjct: 399 IERLFGTAKYHNLRYTREKGSK 422

45 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2865-

Example 2908

A DNA sequence (GASx1990R) was identified in *S.pyogenes* <SEQ ID 8315> which encodes the amino acid sequence <SEQ ID 8316>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2909

A DNA sequence (GASx1991) was identified in *S.pyogenes* <SEQ ID 8317> which encodes the amino acid sequence <SEQ ID 8318>. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -0.16 Transmembrane 2 - 18 (1 - 18)

----- Final Results -----

bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2910

A DNA sequence (GASx1994) was identified in *S.pyogenes* <SEQ ID 8319> which encodes the amino acid sequence <SEQ ID 8320>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.44 Transmembrane 28 - 44 (28 - 44)

----- Final Results -----

bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2866-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2911

A DNA sequence (GASx1996) was identified in *S.pyogenes* <SEQ ID 8321> which encodes the amino acid sequence <SEQ ID 8322>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1076(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2912

A DNA sequence (GASx1997R) was identified in *S.pyogenes* <SEQ ID 8323> which encodes the amino acid sequence <SEQ ID 8324>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -7.96 Transmembrane 53 - 69 (49 - 75)
INTEGRAL Likelihood = -2.34 Transmembrane 24 - 40 (24 - 43)

----- Final Results -----

bacterial membrane --- Certainty=0.4185(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2913

A DNA sequence (GASx2007R) was identified in *S.pyogenes* <SEQ ID 8325> which encodes the amino acid sequence <SEQ ID 8326>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -6.64 Transmembrane 46 - 62 (43 - 65)

----- Final Results -----

bacterial membrane --- Certainty=0.3654(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

-2867-

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB97959 GB:U96166 ATP-binding cassette lipoprotein
[Streptococcus cristatus]
Identities = 37/60 (61%), Positives = 42/60 (69%), Gaps = 1/60 (1%)

Query: 59  FLTACGTTKKDSKKEEVKEIKMSDIKDDAVSKKTKVVDGEEVTEYTTKDGNNVIQIPAGNEE 118
          FL ACG+K   KE + + K  D K DAV +KTK VDG+EVTEYT DGNVIQIPA EE
Sbjct: 12  FLAACGSKNADNKE-ISDGKKVDFKKDAVDQKTKTVDGKEVTEYTMPDGNVIQIPADGEE 70
```

- 10 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2914

A DNA sequence (GASx2009) was identified in *S.pyogenes* <SEQ ID 8327> which encodes the amino acid sequence <SEQ ID 8328>. Analysis of this protein sequence reveals the following:

```
15 Possible site: 41

    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
20          bacterial cytoplasm --- Certainty=0.1246(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

- 25 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2915

A DNA sequence (GASx2010) was identified in *S.pyogenes* <SEQ ID 8329> which encodes the amino acid sequence <SEQ ID 8330>. Analysis of this protein sequence reveals the following:

```
30 Possible site: 17

    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
35          bacterial cytoplasm --- Certainty=0.2549(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

- 40 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2916

- 45 A DNA sequence (GASx2012R) was identified in *S.pyogenes* <SEQ ID 8331> which encodes the amino acid sequence <SEQ ID 8332>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
```

-2868-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3307(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

10 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA27007 GB:L26141 pyrogenic exotoxin B [*Streptococcus pyogenes*]
 Identities = 40/102 (39%), Positives = 57/102 (55%), Gaps = 7/102 (6%)
 Query: 2 EMHFVRTEPEARRIAETFC AENTQTKTPMRVQQLSYPSDTHSGGEL-----YIYALSPA 56
 15 + +F R E EA+ A TF ++ K R + D + GGEL YIY +S
 Sbjct: 28 DQNFARNEKEAKDSAITFIQKSAAIKAGARSAE-DIKLDKVNLG GELSGSNMYIYNISTG 86
 Query: 57 GFIIVSGDTRAHTILGYSFDNNLDLN-HDNVRSMIEAYQKQI 97
 GF+IVSGD R+ ILGYS + D+N +N+ S +E+Y +QI
 20 Sbjct: 87 GFVIVSGDKRSPEILGYSTSGSFDVNGKENIASFMESYVEQI 128

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2917

25 A DNA sequence (GASx2013R) was identified in *S.pyogenes* <SEQ ID 8333> which encodes the amino acid sequence <SEQ ID 8334>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

30 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
 35

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2918

40 A DNA sequence (GASx2014R) was identified in *S.pyogenes* <SEQ ID 8335> which encodes the amino acid sequence <SEQ ID 8336>. Analysis of this protein sequence reveals the following:

Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.1392(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 50

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2869-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2919

A DNA sequence (GASx2015) was identified in *S.pyogenes* <SEQ ID 8337> which encodes the amino acid sequence <SEQ ID 8338>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -1.75 Transmembrane 18 - 34 (17 - 37)

----- Final Results -----

bacterial membrane --- Certainty=0.1702(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2920

A DNA sequence (GASx2018) was identified in *S.pyogenes* <SEQ ID 8339> which encodes the amino acid sequence <SEQ ID 8340>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -5.84 Transmembrane 23 - 39 (22 - 40)

----- Final Results -----

bacterial membrane --- Certainty=0.3336(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2921

A DNA sequence (GASx2019) was identified in *S.pyogenes* <SEQ ID 8341> which encodes the amino acid sequence <SEQ ID 8342>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0669(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2870-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC98898 GB:AF023179 low temperature requirement C protein
[Listeria monocytogenes]

Identities = 95/144 (65%), Positives = 117/144 (80%)

Query: 15 LAERGVSLAIAELVLFQNDYIPNLTMACLESVEAVLAKREVQNAIITGVELDKLAEA 74
L ERGV ++ IAEVLVFLQ Y P L + C ++VE VL KREVQNA++TG++LD +AE

Sbjct: 16 LIERGVEIDDI AELVLFQKKYHPGLELDICRQNVHVLKREVQNAVLTGIQLDVMMAEK 75

Query: 75 NQLSEPLLSILKTDQGLYGIDEILALSIVNLYGSIGFTNYGYLDKTKPGIVDKLNHKDGY 134
+L +PL +I+ D+GLYG+DEILALSIVN+YGSIGFTNYGY+DK KPGI+ KLN DG

Sbjct: 76 GELVQPLQNIISADEGLYGVD EILALSIVNVYGSIGFTNYGYIDKVKPGILAKLNEHDGI 135

Query: 135 SHTFLDDIVSAIAAAAAASRIAHN 158
+ HTFLDDIV AIAAAAAASR+AH+

Sbjct: 136 AVHTFLDDIVGAIAAAAAASRLAHS 159

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2922

A DNA sequence (GASx2030) was identified in *S.pyogenes* <SEQ ID 8343> which encodes the amino acid sequence <SEQ ID 8344>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0320(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2923

A DNA sequence (GASx2031) was identified in *S.pyogenes* <SEQ ID 8345> which encodes the amino acid sequence <SEQ ID 8346>. Analysis of this protein sequence reveals the following:

Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0583(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2871-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2924

A DNA sequence (GASx2032R) was identified in *S.pyogenes* <SEQ ID 8347> which encodes the amino acid sequence <SEQ ID 8348>. Analysis of this protein sequence reveals the following:

Possible site: 53

```
>>> Seems to have no N-terminal signal sequence
```

INTEGRAL Likelihood = -2.76 Transmembrane 27 - 43 (26 - 43)

----- Final Results -----

bacterial membrane --- Certainty=0.2105(Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8467> and protein <SEQ ID 8468> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10

McG: Discrim Score: -11.19

GvH: Signal Score (-7.5): -4.94

Possible site: 49

```
>>> Seems to have no N-terminal signal sequence
```

```
ALOM program      count: 1 value: -4.19 threshold: 0.0
```

INTEGRAL	Likelihood = -4.19	Transmembrane	25 - 41 (25 - 42)
----------	--------------------	---------------	--------------------

PERIPHERAL Likelihood = 13.26 41

modified ALOM score: 1.34

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.2678(Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF01616 (304 - 429 of 771)

SP|006442|SECE STAAU(7 - 48 of 60) PREPROTEIN TRANSLOCASE SECE SUBUNIT.

GP|2078376|qb|AAB54017.1|U96619 SecE {Staphylococcus aureus}

```
%Match = 5.4
```

%Identity = 26.2 %Similarity = 57.1

Matches = 11 Mismatches = 18 Conservative Sub.s = 13

99 129 159 189 219 249 279 309

RIIOTMLK*HLWRRYGTKESKPSVYRMRKPKLLNRSK*HPOANTTRSK*IL*IL*EVYNTORNALI*BNKLOKGEIIMFV

1

MAKKESEF

339 369 399 429 459 489 519 549

KGIFQVLRDITTPWPNRKORWKDFISILEYTVFFETIVTYIFDKLLAAGVMDLINRF***IILDRNNPNP*IILRVFCVENNI

|| : || : || : || : || : || : || : || : || :

KGVKSEMEKTSWPTKEELFKYTVIVVSTVIFFLVFFYALDLGITALKNLLFG

20 30 40 50 60

SEQ ID 8468 (GBS396) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 9; MW 35kDa).

GBS396-GST was purified as shown in Figure 217, lane 8.

5 Example 2925

A DNA sequence (GASx2034R) was identified in *S.pyogenes* <SEQ ID 8349> which encodes the amino acid sequence <SEQ ID 8350>. Analysis of the protein sequence reveals the following:

Possible site: 21

```

10  >>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -0.59      Transmembrane      53 - 69 ( 53 - 70)

      ----- Final Results -----
15          bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

20 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2926

A DNA sequence (GASx2035) was identified in *S.pyogenes* <SEQ ID 8351> which encodes the amino acid sequence <SEQ ID 8352>. Analysis of this protein sequence reveals the following:

```

25  Possible site: 39

      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
30          bacterial cytoplasm --- Certainty=0.2928(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

35 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2927

40 A DNA sequence (GASx2042R) was identified in *S.pyogenes* <SEQ ID 8353> which encodes the amino acid sequence <SEQ ID 8354>. Analysis of this protein sequence reveals the following:

Possible site: 44

```

      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
45          bacterial cytoplasm --- Certainty=0.2547(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

```

-2873-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2928

A DNA sequence (GASx2043) was identified in *S.pyogenes* <SEQ ID 8355> which encodes the amino acid sequence <SEQ ID 8356>. Analysis of this protein sequence reveals the following:

10 Possible site: 26

>>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----

bacterial cytoplasm ---	Certainty=0.3289(Affirmative)	< succ>
bacterial membrane ---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside ---	Certainty=0.0000(Not Clear)	< succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

- 20 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2929

A DNA sequence (GASx2049) was identified in *S.pyogenes* <SEQ ID 8357> which encodes the amino acid sequence <SEQ ID 8358>. Analysis of this protein sequence reveals the following:

25 Possible site: 13

>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----

bacterial cytoplasm ---	Certainty=0.4014(Affirmative)	< succ>
bacterial membrane ---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside ---	Certainty=0.0000(Not Clear)	< succ>

- 35 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2930

- 40 A DNA sequence (GASx2052) was identified in *S.pyogenes* <SEQ ID 8359> which encodes the amino acid sequence <SEQ ID 8360>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have a cleavable N-term signal seq.

45 ----- Final Results -----

-2874-

```

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

5 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2931

10 A DNA sequence (GASx2055R) was identified in *S.pyogenes* <SEQ ID 8361> which encodes the amino acid sequence <SEQ ID 8362>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.3048(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:BAB05703 GB:AP001513 imidazolonepropionase
(imidazolone-5-propionate hydrolase) [Bacillus halodurans]
25 Identities = 203/416 (48%), Positives = 278/416 (66%), Gaps = 4/416 (0%)

Query: 11  DVLLTHFNQLFCLNDPGHPLTGQEMKKATIVEDGYIAIKDGLIVALGSGEPPDAELVGTQT 70
D LL + QL + G P G+EM + ++E + I+DG + +G+ Q
Sbjct: 6  DTLIVNIGQLLPMESKG-PKRKGEMSELQLEHAALGIRDGKVAFIGTMVEADTFANQM 64

Query: 71  IMRSYKGKIATPGIIDCHTHLVYGGGREHEFAKKLAGVSYLDILAQQGGILSTVRATRSA 130
I +GK+ TPG++D HTHL++GGSREHE A K GV YL+IL GGGIL+TV ATR+A
Sbjct: 65  I--DCQGKLVTPGLVDPHTHLIFGGSREHEMALKQQGVYPYLEILKNGGGILATVEATRAA 122

Query: 131 SFDNLYQKSKRLLDYMLLHGVTTVEAKSGYGLDWETEKRLDVVALEKDHPIDLVSTFM 190
S + L K+ L+ ML +GVTT+EAKSGYGLD ETE +QL A+ + HPID+VSTF+
Sbjct: 123 SEELITKAICHNLNRLSYGVTTIEAKSGYGLDRETEWKQLRAAKAVGEQHPIDIVSTFL 182

Query: 191 AAHAIPPEEYKGNPKAYLDVLIKDMLPVVKEENLAEFCDIFCEKNVFTADESRYLKAKE 250
AHAIP ++ +P +LD + DML +KE+NLAEF DIF E VFT ++SR L KAKE
Sbjct: 183 GAHAIPTSHRNDPDRFLDEMA-DMLGEIKEQNLAEFVDIFTETGVFTVEQSRTFLQKAKE 241

Query: 251 MGFKLRIHADEIASIGGVDVAAELSAVSAEHLMMITDDGIAGLIGAGVIGNLLPATTFSL 310
GF L++HADEI +GG ++A EL A+SA+HL+ +D GI K+ AG I LLP TTF L
Sbjct: 242 RGFGKLKHADEIDPLGGAELAGELGAISADHLVGASDQGIQKMAAAGTIACLLPGTTFYL 301

Query: 311 MEDTYAPARKMIDAGMAITLSTDSNPGSCPTANMQFVMQLGCFMLRLTPIEVLNAVITINA 370
+DTYA AR MID G+A+T+STD NPGS PT N+Q +M + L++TP E+ +AVT+N
Sbjct: 302 GKDTYARARDMIDQGLAVTISTDFNPGSSPTENLQLIMSIAALRLKMTPEEIIHHAHTVNG 361

Query: 371 AYSVNRQERVGSLTVGKEADIAIFDAPNIDYPPYFFATNLIHQVYKKGQLTVDRGR 426
A+++ R + G L VG+ AD+ ++DA N Y Y + N +H V+KKG++ +R R
Sbjct: 362 AHAIGRGDTAGQLAVGRAADVVDKKNYYYVPYHYGVNVHSHVWKKGEVVYERRR 417

```

55 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2875-

Example 2932

A DNA sequence (GASx2056) was identified in *S.pyogenes* <SEQ ID 8363> which encodes the amino acid sequence <SEQ ID 8364>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1847(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB61139 GB:AL132952 predicted using Genefinder~cDNA EST
yk155e6.3 comes from this gene~cDNA EST yk155e6.5 comes
from this gene~cDNA EST yk156d6.5 comes from this
gene~cDNA EST yk259b10.3 comes fr
Identities = 302/649 (46%), Positives = 419/649 (64%), Gaps = 17/649 (2%)

Query: 29 EGIRRAPDRGFRRLTQAQTEIALKNALRYVPTKFHEEVIPEFLEELKTRGRIYGYRFRPKD 88
++ AP R LTQ + +A++NALRY+P + H + EF EEL T G IYGYR F P
Sbjct: 85 KNVAHAPKRPCNLQTQTEKMLAVRNALRYIPKEHHVLLATEFAEELNTYGHYGYRFPMPNF 144

Query: 89 RIYKGPIDYKGNCTAAKAMQVMIDNNLSFEIALYPYELVTYGETGSVCANWMQYCLIKK 148
++ P+ E +C A A+ +MI NNL +A +P ELVTYG G V +NW+Q+ L+ +
Sbjct: 145 DLFAPPVSEIGAHCEQASAILMILNNLDKRVAQFPQELVTYGGNGQVFSNWIQFRLVLR 204

Query: 149 YLEVMTDEQTLVVESGHPVGLFKSKPEAPRVIIITNGLLVGEYDNMKDWEIAEEMGVTNYG 208
YL MTD QTLV+ SGHP+GLF S P++PR+ +TNG+++ Y + ++ +GVT YG
Sbjct: 205 YLYTMDTHQTLVLYSGHPLGLFPSTPDSPRMTVTNGMMIPSYSTKELYDKYFALGVTQYG 264

Query: 209 QMTAGGWMYIGPQGI VHGTFTNTLLNAGRLKLG VADDDLTGKLFISSGLGMSGAGQKAA 268
QMTAG + YIGPQGI VHGT T+LNAGR ++G+ L GK+F+++GLGMSGAG QAA
Sbjct: 265 QMTAGSFCYIGPQGI VHGTITITVLNAGR-RMGL---DSLAKGVFVTAGLGMSGAGQPKAA 320

Query: 269 EIAKAVAIIEVDQSRIKTRHSQGWISQIAESPEEALQLAQKAIDAKESTSIAYHGNIVD 328
+IA + +IAE+ + + RH QGW+ ++ EE + ++ + KE+ SI Y GN+VD
Sbjct: 321 KIAGCIGVIAEISDTALLKRHQGWLDVYSKDLBEIVNWIKEYREKKEAISIGYLGNNVD 380

Query: 329 LLE-YVNDKQIHVDLLSDQTSCHNVYDGGYCPVGISFDERTLLAEDKDTFHMVDDTLA 387
L E + + V+L SDQTS HN + GG+ P G++F++ +++ D F ++V ++L
Sbjct: 381 LWERLAAEEPECLVELGSDQTSLHNPFLGGFYPAGLTFEQSNQMMTSDPVKFKLVQNSLI 440

Query: 388 RHFEAIKTLTENGTYFFDYGNAFMKSVDYSGITEISKNGRNDKDGFIWPSYVEDIMGPML 447
R AI + G YF+DYGNAF+ +G + ++ ++DK F +PSY++DIMG +
Sbjct: 441 RQIAAIDKIAAKGMYFWDYGNAFLLLECQRAGANLLREDAQDDK-SFRYPSYMQDIMGD-I 498

Query: 448 FDYGYGPFRWVCLSGNHDDLVAATKAAMEAIDPDR-----RYQDRDNYNWIRDAEKN 499
F G+GPFRWVC SG +DL TD+ A + ID + + Q DN WI +AEKN
Sbjct: 499 FSMGFGPFRWVCTSGKPEDLRLTDQTACKI IDELKDTDVPEYVKQYLDNKKWIEEAEKN 558

Query: 500 QLVVGTQARILYQDCIGRVTIALKFNELVKGGKI-GPVMIGRDHHDVSGTDSPPFRETSTNI 558
+LVVG+QARILY D GRV +A FNELV+ GK+ ++I RDHHDVSGTDSPPFRETSTN+
Sbjct: 559 KLVVGSQARILYSDRAGRVALASAFNELVKSGKVSAAIVISRHDHHDVSGTDSPPFRETSTNV 618

Query: 559 KDGSNVTCDMAVQCYAGNAARGMSLVALHNGGGTGIGKAINGGFGVLVDGSRIDEIIS 618
DGS T DMAVQ G++ RG + VALHNGGG G G INGGFG+VLDGS +
Sbjct: 619 YDGSFTADMAVQNCIGDSFRGATWVALHNGGGVGWGDVINGGFGIVLDGSSDAARRAEG 678

Query: 619 AIAWDTMGGVARRNWARNEHAIEAIEYNRLHAGTDHITIPYLADDDL 667
+ WD GV RR+W+ N A E AI+ +T+P AD++L+
Sbjct: 679 MLNWDVPNGVIRRSWSGNAKAE-AIQRAEKQVDGLRVTLFVEADEELL 726

-2876-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2933

- 5 A DNA sequence (GASx2057) was identified in *S.pyogenes* <SEQ ID 8365> which encodes the amino acid sequence <SEQ ID 8366>. Analysis of this protein sequence reveals the following:

Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1887(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD35925 GB:AE001751

formiminotransferase-
cyclodeaminase/formiminotetrahydrofolate cyclodeaminase,
putative [Thermotoga maritima]

Identities = 160/296 (54%), Positives = 214/296 (72%), Gaps = 2/296 (0%)

Query: 3 KIVECIPNFSEGNQAVIDGLVATAKSIPGVTLDDYSSDASHNRSVFTLVGDDQSIQEAA 62

K++E +PNFSEG+ + V++ +VA AK V +LD+S DA HNRSV TLVG+ +++ A

Sbjct: 2 KLIESVPNFSEGRRKEVVEKIVAEAKKYDRVWVLDWSMDADHNRSVITLVGEPENLINAL 61

Query: 63 FQLVKYASENIDMTKHHGEHPRMGATDVCPFVPIKDITTQECVEISKQVAERINRELGIP 122

F + K A+E ID+ H G+HPRMGA DV P VP+ + T +ECVE SK + RI ELGIP

Sbjct: 62 FDMTKKAELIDLNRNHTGQHPRMGAADVIPLVPLYNITMEECVEYSKILGRRIGEELGIP 121

Query: 123 IFLYEDSATRPERQNLAKVRKQGFEGMPEKLLEEDWAPDYGDRKIHPTAGVTAVGARMPL 182

++LYE SATRPERQNLA +RKG+FEG EK+ + W PD+G ++HPTAGVTAVGAR L

Sbjct: 122 VVLYEKSATRPERQNLADIRKGEFEGFFFEKIKDPLWKPDFGPDVRHPTAGVTAVGAREFL 181

Query: 183 VAFNVNLDTDNIDIAHKIAKIIRSGGGYKYCKAIGVMLEDRHIAQVSMNMVNFECSLY 242

+AFNVNL T ++ IA KIA+ IR S GG +Y KAIGV L+ R + QVS+N+ N +K LY

Sbjct: 182 IAFNVNLGTRDVKIAEKIARAIRFSSGGLRYVKAIGVDLKGGRVVQVSINITNHHKTPLY 241

Query: 243 RTFETIKFEARRYGVNVIGSEVIGLAPAKALIDVAEYYLQVEDFDYHKQILENHLL 298

R FE IK EA RYGV V+GSE++GL P ++L+ YYL+ + K+++E++LL

Sbjct: 242 RVFELIKMEAERYGVFVLGSEIVGLFPLESLKTVSYLRTD--LNAAKVIESNLL 295

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2934

A DNA sequence (GASx2058) was identified in *S.pyogenes* <SEQ ID 8367> which encodes the amino acid sequence <SEQ ID 8368>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2776(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2877-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

5  >GP:AAA62653 GB:L33465 methenyl tetrahydrofolate cyclohydrolase
    [Methylobacterium extorquens]
    Identities = 79/198 (39%), Positives = 112/198 (55%)

Query: 7  SLTDFAKVLGSDAPAPGGGSAALSGANGISLTKMVCELTLGKKKYADYQDIITEIHAKS 66
      ++  F   L S AP PGGG AAA+SGA G +L MVC LT+GKKKY + + + ++ KS
10  Sbjct: 6  TIETFLDGLASSAPTPGGGAAAISGAMGAALVSMVCNLTIGKKKYVEEADLMQVLEKS 65

Query: 67  TALQASLLAAIDKDTEAFNLVSAVFDMPKETDEDKAARRTAMQKALKTAAQSPFEMMTLM 126
      L+ +L   I D EAF+ V   +PK TDE+KAAR +Q+ALKTA P   +
15  Sbjct: 66  EGLRRTLTLGMIADDVEAFDAVMGAYGLPKNTDEEKAARAQKIQEALKTATDVPLACCRVC 125

Query: 127  VEALEITATAVGKSNTINAASDLGVAALNLKAGLQGAWLNLINLSGIKDEDFVTDYRQKG 186
      E +++      K N N SD GVA L+  AGL+ A LNV +N G+ D F + ++
20  Sbjct: 126  REVIDLAEIVAEKGNLNVISDAGVAVL SAYAGLRSALNVYVNAKGLDDRAFAEERLKEI 185

Query: 187  QALLDKGCHLADDIYTKI 204
      + LL +   L + IY +
20  Sbjct: 186  EGLLAEAGALNERIYETV 203

```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 25 antigens for vaccines or diagnostics.

Example 2935

A DNA sequence (GASx2061) was identified in *S.pyogenes* <SEQ ID 8369> which encodes the amino acid
 sequence <SEQ ID 8370>. Analysis of this protein sequence reveals the following:

```

30  Possible site: 22

    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
35      bacterial cytoplasm --- Certainty=0.3924 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

40 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 antigens for vaccines or diagnostics.

Example 2936

A DNA sequence (GASx2063) was identified in *S.pyogenes* <SEQ ID 8371> which encodes the amino acid
 sequence <SEQ ID 8372>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 57

    >>> Seems to have an uncleavable N-term signal seq
        INTEGRAL    Likelihood = -1.06    Transmembrane  231 - 247 ( 231 - 247)
50  INTEGRAL    Likelihood = -0.53    Transmembrane    2 - 18 ( 1 - 18)

    ----- Final Results -----
        bacterial membrane --- Certainty=0.1426 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

-2878-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:CAB15971 GB:Z99124 histidase [Bacillus subtilis]
Identities = 236/477 (49%), Positives = 321/477 (66%), Gaps = 2/477 (0%)

Query: 42 VINLDGESLTIEDVIAIARQGVACHIDDSAIEAVNASRKIVDDIVSEKRVVYGVITGFGS 101
++ LDG SLT DV + + ++E V SR V+ IV +++ +YG+ TGFG

10 Sbjct: 1 MVTLDGSSLTITADVARVLFDFEEAAASESMERVKKSRAAVERIVRDEKTIYGINITGFGK 60

Query: 102 LCNVISISPEDTVQLQENLIRTHASGFGDPLPEDAVRAIMLIRINSLVKGYSGIRLSTIEK 161
+V I ED+ LQ NLI +HA G GDP PE RA++L+R N+L+KG+SG+R IE+

15 Sbjct: 61 FSDVLIQKEDSAALQLNLILSHACGVGDPFPECVSRAMLLLRANALLKGFSGVRAELIEQ 120

Query: 162 LLELLNKGVHPYIPEKGSIGASGDLAPLAHMVLPMLGLGKAYYKGELLSGQEAIDKAGID 221
LL LNK VHP IP++GSLGASGDLAPL+H+ L ++G G+ +++GE + L KAGI

Sbjct: 121 LLAPLNKRVPVPIQQGSLGASGDLAPLSHLALALIGQGEVFFEGERPAMTGLKKAGIQ 180

20 Query: 222 KISLAAKEGLALINGTTVLTAVALATYDAIQLLKLSDLAGALSLEVHNGITSPPFEENLH 281
++L +KEGLALINGT +TA+G +A +A +L ++ +L++E GI F+E++H

Sbjct: 181 PVTILTSKEGLALINGTQAMTAMGVVAYIEAEKLAYQTERIASLTIEGLQGIIDAFDEDIH 240

25 Query: 282 TIRPQSGQLATARNIRNLLGSSQNTTVATQSRVQDPYTLRCMPQIHGASKDSIAYVKSIV 341
R Q+ A IR L S TT + RVQD Y+LRC+PQ+HGA+ ++ YVK K+

Sbjct: 241 LARGYQEQIDVAERIRFYLSDSGLTTSQGELRVQDAYSLRCIPQVHGATWQTLGYVKEKL 300

Query: 342 DIEINSVTDNPIICKDG-HVISGNFHFGEPMAPDFDLGIAISEIGNVSERRVERLVNSQ 400
+IE+N+ TDNP+I DG VISGNFHF+P+A DFL IASE+ N++ERR+ERLVN Q

30 Sbjct: 301 EIEMNAATDNPLIFNDGDKVISGNFHFHQPIAFAMDFLKIAISELANIAERRIERLVNPQ 360

Query: 401 LSKLPSPFLVKYPGLNSGFMITQYACASLASSENKVLAPASVDSIPSCENQEDFVSMGTTA 460
L+ LP FL +PGL SG MI QYA ASL SENK LAHPASVDSIPS NQED VSMGT A

35 Sbjct: 361 LNDLPPFLSPHPGLQSGAMIMQYAAASLVSENKTLAHPASVDSIPSSANQEDHVSMTIA 420

Query: 461 ARKAFFELKNSRRIVATEIMAAQALDLKPENHELKGKTKVAYDLFRKEVNFIEHDK 517
AR A++++ N+RR++A E + A QA++ + H TK + RK V I+ D+

Sbjct: 421 ARHAYQVIANTRRRIAIEAICALQAVEYRGIEH-AASYTKQLFQEMRKVVPSIQQDR 476

40 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2937

A DNA sequence (GASx2064) was identified in *S.pyogenes* <SEQ ID 8373> which encodes the amino acid sequence <SEQ ID 8374>. Analysis of this protein sequence reveals the following:

45 Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.4483(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

55 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG06563 GB:AE004741 probable arginase family protein
[Pseudomonas aeruginosa]
Identities = 99/275 (36%), Positives = 147/275 (53%), Gaps = 9/275 (3%)

-2879-

Query: 53 LIGFKSDKGVYINNGRVGAVESPAAIRTQLAKFPWHLGNQVMVYDVGNIDGPNRSLEQLQ 112
 L+GF SD+GV N GR GA P A+R LA WH G Q +YD G+I + LE Q
 Sbjct: 42 LLGFASDEGVRRNQGRQGARHGPPALRRALANLAWH-GEQA-TYDAGDITVAGD-DLEAAQ 98

Query: 113 NSLSKAIKRMCDLNLKPIVLGGGHETAYGHYLGRLQSLSPSDDL--AVINMDAHFDLRP 169
 ++ + + + LGGGHE AY + GL + LS + L ++N DAHFDLR
 Sbjct: 99 ECYAQRVADLLACGHRVVGGLGGGHEIAYASFAGLARHLSRHERLPRIGILNFDHFDLRH 158

Query: 170 YDQTGPNSTGTGFRQMFDDAVADKRLFKYFVLGIQEHNNLFLFDVFAKSKGIQFLTQODI 229
 ++ +SGT FRQ+ + A F Y LGI +N LFD A+ G+++L + +
 Sbjct: 159 AERA--SSGTPFRQIAELCQASDWPFAYCCCLGISRLSNTAALFD-QAQLGVRYLLDRQL 215

Query: 230 YQMGHQKVCRAIDRFLEGQERVYLTIDMDCFSVGAAPGVSAIQSLGVDPNLAVLVLQHIA 289
 ++ +D FL+ + +YLT+ +D APGVSA + GV+ + +++
 Sbjct: 216 QPWNLERSEAFLDGFLQSVDHLYLTVCCLDVLPAQAAPGVSAAPSAGHVEMPVVEHLVRRAK 275

Query: 290 ASGKLVGFDVVEVSPPHDIDNHTANLAATFIFYLV 324
 ASGKL D+ E++P D D TA +AA + LV
 Sbjct: 276 ASGKLRLADIAELNPQLDSQRTARIAARLVDSL 310

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2938

A DNA sequence (GASx2065R) was identified in *S.pyogenes* <SEQ ID 8375> which encodes the amino acid sequence <SEQ ID 8376>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.37 Transmembrane 375 - 391 (375 - 392)

----- Final Results -----

bacterial membrane --- Certainty=0.1150(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB37582 GB:AL035569 putative regulatory protein [Streptomyces
 coelicolor A3(2)]
 Identities = 95/437 (21%), Positives = 177/437 (39%), Gaps = 28/437 (6%)

Query: 271 EVGALLLIGDTGIGKRTLARQVLANQTQTFQIVTAKCFREEAMDSL--LPWRNILDGLGD 328
 E ALLL G+ G+GK L + A + +V E D L P+ L L
 Sbjct: 95 EPQALLLGGEAGVGKTRLVEEFAAADRRGAVVALGGCVEIGADGLPFAPFSTALRALRR 154

Query: 329 LVIQNRLLTTKAWKAALKRCFP-VATIFQEDNNQPFIKDHTSLVSVFIVDILQHLAEIKA 387
 + + + L R P +A ++ + L +L+ +A
 Sbjct: 155 HLPEELAAAAAQEEELARLLPELAEGTPVTGGGRHDEESMARLFELTARLLERVAARHT 214

Query: 388 LVILIEDCHWMDSDLTLLQRVMNQIVHYPIAFVLT-----KHLGTTPELGCLCLNALM 440
 +V+++ED HW D + L+ ++ L + + T + P L L+ L
 Sbjct: 215 VVLVLEDLHWADASTRHLLIAYLLRLTTLTGRLVVLATYRSDDIHRRLPLRLAE-LDRLR 273

Query: 441 SQGRLESICLEPFNRQESLVYINSQIGSQPVTAEMEHLQASQGNPFFLSEYTOALLRH 500
 + RLE L F R E I L +P +++ +++ S GN FF+ E A R
 Sbjct: 274 TVRRLE--LGRFTRDEVGRQIAGILAHEP-DQLQVDEIFERSDGNAFFVEELAVA-ARV 328

Query: 501 EKFPVPLTPAIKAKLGLKLANLSSRDDALLNYLSCRRPIPLNTLAQLMLLPLEEVIEMVD 560
 LT +++ L +++ L + ++ + LA + L +++IE +
 Sbjct: 329 GSCTGLTDSLRLDLLVRVEALPESAQRVARIVAEGGSTVEYRLLAARLAEDDLIEALR 388

-2880-

Query: 561 NLGHYYILVEESVGEEVLISFRQRIIQLYSYDRLSLSKRRLHGGQIAKRIEDLLPILTPS 620
 + + IL+ G+ FR +++ D L +R L+ + A+ L D P L P+
 5 Sbjct: 389 SAVNANILLPAPDGDG--YRFRHSLVREAVGDDLLPGER SRLNRRYAEAL-DADPTLVPA 445

Query: 621 PHLLDDIAYHYQESRQVIKALEYNLNLYLDATLPFOHELFPIYSKSGSLEKSDRDHQRIM 680
 + +A ++ + KAL LDA++ + YS+ + LE++ L
 10 Sbjct: 446 AERMRLASYWYHAHAPAKALP---AVLDASVEARRR--HAYSEQLRLLEA----MELW 496

Query: 681 EEQFDKIRQSIADLELT 697
 + D +R ++ ++ T
 15 Sbjct: 497 DSAPDDVRATLRPVDC 513

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 15 antigens for vaccines or diagnostics.

Example 2939

A DNA sequence (GASx2072) was identified in *S.pyogenes* <SEQ ID 8377> which encodes the amino acid
 sequence <SEQ ID 8378>. Analysis of this protein sequence reveals the following:

Possible site: 14
 20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.3702(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 30 antigens for vaccines or diagnostics.

Example 2940

A DNA sequence (GASx2074R) was identified in *S.pyogenes* <SEQ ID 8379> which encodes the amino
 acid sequence <SEQ ID 8380>. Analysis of this protein sequence reveals the following:

Possible site: 37
 35 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.90 Transmembrane 21 - 37 (21 - 38)

----- Final Results -----
 40 bacterial membrane --- Certainty=0.1362(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 antigens for vaccines or diagnostics.

-2881-

Example 2941

A DNA sequence (GASx2075R) was identified in *S.pyogenes* <SEQ ID 8381> which encodes the amino acid sequence <SEQ ID 8382>. Analysis of this protein sequence reveals the following:

Possible site: 25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```
bacterial cytoplasm --- Certainty=0.3545(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2942

A DNA sequence (GASx2076R) was identified in *S.pyogenes* <SEQ ID 8383> which encodes the amino acid sequence <SEQ ID 8384>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```
bacterial cytoplasm --- Certainty=0.2340(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC44494 GB:U44893 orf108; unknown function [Butyrivibrio
fibrisolvens]
Identities = 42/75 (56%), Positives = 55/75 (73%)
```

```
Query: 1 LLKGTLRFGQLKSSIGSVSQKVLTAQLRAMEADGLVHREVYAEPVPPRVEYSLTETGLSLA 60
LL RF +LK+++ +SQKVL LR+ME DG++ R VY EVPPRVEYSL+E G S+
Sbjct: 31 LLVRPWRFNELKNNLEGISQKVLTDLSRMEEDGIITRTVYPEVPPRVEYSLSELGESMR 90
```

```
Query: 61 PVIEAMSDWGQTYQE 75
P+I+AM WG Y+E
Sbjct: 91 PIIKAMEQWGTEYKE 105
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2943

A DNA sequence (GASx2097) was identified in *S.pyogenes* <SEQ ID 8385> which encodes the amino acid sequence <SEQ ID 8386>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

```
INTEGRAL Likelihood = -3.40 Transmembrane 26 - 42 ( 23 - 44)
```

-2882-

----- Final Results -----

5 bacterial membrane --- Certainty=0.2359(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

10 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 antigens for vaccines or diagnostics.

Example 2944

A DNA sequence (GASx2098) was identified in *S.pyogenes* <SEQ ID 8387> which encodes the amino acid
 sequence <SEQ ID 8388>. Analysis of this protein sequence reveals the following:

15 Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.1385(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

25 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 antigens for vaccines or diagnostics.

Example 2945

A DNA sequence (GASx2100) was identified in *S.pyogenes* <SEQ ID 8389> which encodes the amino acid
 sequence <SEQ ID 8390>. Analysis of this protein sequence reveals the following:

30 Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.2138(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

40 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA98589 GB:L44593 ORF79; putative [Lactococcus lactis phage
 BK5-T]

Identities = 34/62 (54%), Positives = 44/62 (70%)

45 Query: 3 QITLKAARINAGYTLKQVAGAVGKNPQTISKYEKDSISLGLLQKLSSLYGVTTIDNLF 62
 +I LKAAR NA ++ K+VA VGKN QTI YEKDS++I + L KL+ +Y ID +FL
 Sbjct: 8 KIKLKAARTNADFSAKEVAEIVGKNYQTILSYEKDSTEIPMSLAIKLAETDYDIPIDFIFL 67

50 Query: 63 GK 64
 GK

Sbjct: 68 GK 69

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2946

5 A DNA sequence (GASx2103) was identified in *S.pyogenes* <SEQ ID 8391> which encodes the amino acid sequence <SEQ ID 8392>. Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3316(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 2947

A DNA sequence (GASx2104) was identified in *S.pyogenes* <SEQ ID 8393> which encodes the amino acid sequence <SEQ ID 8394>. Analysis of this protein sequence reveals the following:

Possible site: 55

25 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4371(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
35 antigens for vaccines or diagnostics.

Example 2948

A DNA sequence (GASx2105) was identified in *S.pyogenes* <SEQ ID 8395> which encodes the amino acid sequence <SEQ ID 8396>. Analysis of this protein sequence reveals the following:

Possible site: 40

40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2263(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

-2884-

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2949

- 5 A DNA sequence (GASx2106) was identified in *S.pyogenes* <SEQ ID 8397> which encodes the amino acid sequence <SEQ ID 8398>. Analysis of this protein sequence reveals the following:

Possible site: 32

```

10 >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -6.42    Transmembrane    9 - 25 ( 6 - 29)

    ----- Final Results -----
                bacterial membrane --- Certainty=0.3569(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

- 20 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2950

A DNA sequence (GASx2107) was identified in *S.pyogenes* <SEQ ID 8399> which encodes the amino acid sequence <SEQ ID 8400>. Analysis of this protein sequence reveals the following:

```

25 Possible site: 25

    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
                bacterial cytoplasm --- Certainty=0.1355(Affirmative) < succ>
30                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

- 35 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2951

A DNA sequence (GASx2108) was identified in *S.pyogenes* <SEQ ID 8401> which encodes the amino acid sequence <SEQ ID 8402>. Analysis of this protein sequence reveals the following:

```

40 Possible site: 26

    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
45                bacterial cytoplasm --- Certainty=0.3050(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-2885-

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
5 antigens for vaccines or diagnostics.

Example 2952

A DNA sequence (GASx2109) was identified in *S.pyogenes* <SEQ ID 8403> which encodes the amino acid
sequence <SEQ ID 8404>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3628(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB46557 GB:AJ242479 putative replication protein [Streptococcus thermophilus]
Identities = 143/242 (59%), Positives = 180/242 (74%), Gaps = 2/242 (0%)

Query: 1 MAIYEARGFSSYLY--PYKGPLEPFDYIAQFRPLKPPEDIDIEEYKRTQAPYCLSGKVTA 58

MAIYE+RGF + L+ +PF ++A FRP+K P+ DI ++KR APYC+SG+V

Sbjct: 1 MAIYESRGFGNHLNNSNASKDPFKFVATFRPMKVPQGEDIADFKRYHAPYICISGEVKQ 60

Query: 59 EKNGSYKRNNASLVYRDLIFLDYDEIETGVNLPKIVSQTLWEYSYIIYPTIKHTPEKPRY 118

+++G+YKRNNASL+YRDLIFLDYD++E + P+ VS L YSY+IYPTIKHT EKPRY

Sbjct: 61 DEDGNYKRNNASLLYRDLIFLDYDKLEASTDFPRAVSNALNGYSYVIYPTIKHTAEKPRY 120

Query: 119 RLVMKPSDVMTEATYKQVVKIEADKIGLPFDLASLTWSQLQGLPVTGDPEDYQRYVNHG 178

RLV+KP+D M E TYK +EADKIGLPFD +SLTWSQLQGLPVTGDP E Y+R VN G

Sbjct: 121 RLVVKPTDKMDEQTYKATAQEIADKIGLPFDDSSLTWSQLQGLPVTGDP EKYERIVNRG 180

Query: 179 LDYPVPKNGSTPNRQVVTYTPRPRSQRISITMRVIDTLFNGFGNEGGRNVALTKFVGLLF 238

YPV + +TPR +S+TMRV+DTL NGFG+EGGRN+ +T+FVGLL

Sbjct: 181 RCYPVANPNTVKANHSPNYHTPRQSGDKSLTMRVVDTLNGFGDEGGRNIEVTRFVGLLL 240

Query: 239 NK 240

+K

Sbjct: 241 SK 242

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
antigens for vaccines or diagnostics.

Example 2953

A DNA sequence (GASx2110) was identified in *S.pyogenes* <SEQ ID 8405> which encodes the amino acid
sequence <SEQ ID 8406>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5215(Affirmative) < succ>

-2886-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S. agalactiae*.

5 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB46558 GB:AJ242479 putative DNA primase [Streptococcus thermophilus]
 Identities = 274/548 (50%), Positives = 363/548 (66%), Gaps = 17/548 (3%)

10 Query: 17 DLKMLENEITEARE-----NEDKYFSTFKGVRGQLIKECQEMKDEAFKIAVDGVMADSK 70
 DL LE E E+++ +ED Y TFK +R Q I ++ K+ A++ YD M + K
 Sbjct: 8 DLTKLEEEYNESKKEASTLFDEDEGYLKTFRKDIRKQFINILEQKKEIAYQKGYDLYMNNPK 67

15 Query: 71 HLENVKAGRLTEVQHE-----ELAKEKGQEAASEKALPKTPLGVAIMLKHYLRFIRVKP 123
 L + E E E AK++G++A + A PKTPL A LK Y+RFIR++P
 Sbjct: 68 VLLKLAKAEKDEENGELIRKTVIEDAKKEGEKAKKNATPKTPLECAEFLKKYIRFIRIRP 127

20 Query: 124 EAQGQKAPLYFFPHDPHGVWLEDNEFLQDLISVIFPNATEKQAFDTLYKIARQSQLKEIQR 183
 + +G++ F G++LED+EFL DL+ I PN TE+ D LYKIA LK+ Q
 Sbjct: 128 KGKGRERLYTFTRQILGIYLEDEFLHDLMTIHPNTERLGNDALYKIAHSVPLKDKQE 187

25 Query: 184 EYTVIGNQLYNYKTGQFEELTPDITVTRKIKTGYNKKAKEPTIKGWKPTAWLLELFDGDA 243
 Y V+G +LYN +TG+F + P I VTRK++ GYN A EP I GWKPT WL LF+GD
 Sbjct: 188 NYVVVGELYNNETGEFTQFDPRIIVTRKVRMGYNPDATPEIIDGWKPTVWLKGLFNGDR 247

30 Query: 244 ELYNLAIQIIKASITGQSLQKIFWLFGEGETGKGTGQQLLINLVGMDNVASLKITELAKS 303
 + Y+LAIQII+A+ITG++L+ IFWL+GEGGTGKGTGQ LL NLVG +NVAS KI + A
 Sbjct: 248 DSYDLAIQIIRATITGKTLENIFWLYGEGGTGKGTGQTLLENLVGSENVASFKEI-DGASG 306

35 Query: 304 RFTTSILLGKSIVIGDDIQDAVIKDTSDIFSLATGDMITIEDKGRPYISIRLNMTVVQS 363
 +F TSIL+GK++VIGDDIQD VIKDTS +FSLATGD + IEDKGRPY+ R MIVVQS
 Sbjct: 307 KFDTSILIGKTVVIGDDIQDVKVVDTSVVFSLATGDPIRIEDKGRPYTTRKRMVVQS 366

40 Query: 364 SNGLPRMNGDKSAIDRRFRILPFTTKVFKGKPKNAIRNDYINRKEVLEYLLKLAIETPITD 423
 SNG PRMN D+ AI+RRFR+L F+++ KGK +K I+NDY+ RKEVLEY +KLAIETP D
 Sbjct: 367 SNGFPRMNADQKAINRRFRVLTFSSEL-KGKADKRIKNDYVGRKEVLEYFVKLAIETPFRD 425

45 Query: 424 INPKASIEILEEHHKEMNPVIDFVSKFFTTDE-LTSEFIPNSFVYHVWKGFLYYDIKQ-I 481
 +NP+ SIE L+E +KEMNPV DFV +FF DE + ++PN +V+ +K + E +
 Sbjct: 426 VNPQKSIEFLDEAYKEMNPVADFVDRFFNDEVIKCNYVPNGYVFECFKAYCEKNQNRNYF 485

50 Query: 482 KSERGLHKEIKSNLPEGFEAGQKVI PVGRQLHTGFYPKEDLPLFASASYANGRASPEKRRK 541
 + R LHK+IK LP+ F + I G++ + F P + +Y NGR E ++
 Sbjct: 486 LNSRTLHKQIKKILPKTFRPKVETIKKGQKFYEEFNPHLVSNPWHPDAYDNGRNKKEDQQ 545

45 Query: 542 KPKNERGY 549
 K ERGY
 Sbjct: 546 DAKKERGY 553

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
 50 antigens for vaccines or diagnostics.

Example 2954

A DNA sequence (GASx2111) was identified in *S. pyogenes* <SEQ ID 8407> which encodes the amino acid
 sequence <SEQ ID 8408>. Analysis of this protein sequence reveals the following:

Possible site: 41

55 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

60 bacterial cytoplasm --- Certainty=0.0994 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
5 antigens for vaccines or diagnostics.

Example 2955

A DNA sequence (GASx2112) was identified in *S.pyogenes* <SEQ ID 8409> which encodes the amino acid
sequence <SEQ ID 8410>. Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3058(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
20 antigens for vaccines or diagnostics.

Example 2956

A DNA sequence (GASx2114) was identified in *S.pyogenes* <SEQ ID 8411> which encodes the amino acid
sequence <SEQ ID 8412>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2815(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful
35 antigens for vaccines or diagnostics.

Example 2957

A DNA sequence (GASx2115R) was identified in *S.pyogenes* <SEQ ID 8413> which encodes the amino
40 acid sequence <SEQ ID 8414>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-2888-

```

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

5 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2958

10 A DNA sequence (GASx2116) was identified in *S.pyogenes* <SEQ ID 8415> which encodes the amino acid sequence <SEQ ID 8416>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
```

```
>>> Seems to have no N-terminal signal sequence
```

15 ----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.4213 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

20 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2959

25 A DNA sequence (GASx2117) was identified in *S.pyogenes* <SEQ ID 8417> which encodes the amino acid sequence <SEQ ID 8418>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
```

```
>>> Seems to have no N-terminal signal sequence
```

30 ----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.3091 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

35

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 2960

A DNA sequence (GASx2118) was identified in *S.pyogenes* <SEQ ID 8419> which encodes the amino acid sequence <SEQ ID 8420>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
```

45 >>> Seems to have an uncleavable N-term signal seq

-2889-

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

10 Example 2961

A DNA sequence (GASx2119) was identified in *S.pyogenes* <SEQ ID 8421> which encodes the amino acid sequence <SEQ ID 8422>. Analysis of this protein sequence reveals the following:

Possible site: 22

15 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2531 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

20

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF63071 GB:AF158600 gp137 [Streptococcus thermophilus
 bacteriophage Sfil11]

25

Identities = 41/121 (33%), Positives = 65/121 (52%), Gaps = 3/121 (2%)

Query: 4 KNAIRKLKEFHRWQRIAN-SLDLTYTELYQFDIEYHPTRR--KHLEISRECALEELDAIR 60

30

Sbjct: 13 K RKL+E+ RW+ IA+ S + T+ + F + +++ + R AL EL+AI

Query: 61 YAINQLSKVEYRQILIECYLISEEKTQDIMEELNGSQSWYYESKKRALLEFVEFYRDGAL 121

A+N L + +YR+ILIE YL K I + + ++ + E ++L F E YRDG L

35

Sbjct: 73 QAVNGLYRPDYRRILIEKYLAYPPKPNWQIAQSIGFERTAFQELLNNSILAFaelYRDGRL 133

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2962

A DNA sequence (GASx2120) was identified in *S.pyogenes* <SEQ ID 8423> which encodes the amino acid sequence <SEQ ID 8424>. Analysis of this protein sequence reveals the following:

40

Possible site: 24

>>> Seems to have no N-terminal signal sequence

45

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2666 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2890-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2963

A DNA sequence (GASx2121) was identified in *S.pyogenes* <SEQ ID 8425> which encodes the amino acid sequence <SEQ ID 8426>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2964

A DNA sequence (GASx2123R) was identified in *S.pyogenes* <SEQ ID 8427> which encodes the amino acid sequence <SEQ ID 8428>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3441(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2965

A DNA sequence (GASx2132) was identified in *S.pyogenes* <SEQ ID 8429> which encodes the amino acid sequence <SEQ ID 8430>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

-2891-

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2966

- 5 A DNA sequence (GASx2136) was identified in *S.pyogenes* <SEQ ID 8431> which encodes the amino acid sequence <SEQ ID 8432>. Analysis of this protein sequence reveals the following:

Possible site: 30

```

10 >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -3.19    Transmembrane    57 - 73 ( 54 - 78)

    ----- Final Results -----
                bacterial membrane --- Certainty=0.2275(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

20 >GP:AAB18271 GB:U74623 CadX [Staphylococcus lugdunensis]
    Identities = 50/110 (45%), Positives = 76/110 (68%)

    Query: 11  MKKDSICQVGVINQQNVTTATNYLEKEKVQKSLRILSKFTDNKQINIIIFYLLAVEELCVC 70
                M  ++ C V  +++  V  A ++LE +K +K L IL K  D K++ II  L+  +ELCVC
    Sbjct: 1    MSYENACDVICVHEDKVNNAISFLEDDKSKLLNILEKICDEKKLKIILSLIKEDELVC 60

25 Query: 71    DIACLINLSMASASHHLRKLANQNILDTRREGKIYYFIKDEEIRDFFNQ 120
                DI+ +L +S+AS SHHLR L  ++LD ++GK+ YYFIKD+EIR+FF++
    Sbjct: 61    DISLILKMSVASTSHHLRLLYKNDVLDIFYKKGKMAYYFIKDEIREFFFSK 110

```

- 30 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2967

A DNA sequence (GASx2137) was identified in *S.pyogenes* <SEQ ID 8433> which encodes the amino acid sequence <SEQ ID 8434>. Analysis of this protein sequence reveals the following:

```

35 Possible site: 49

    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
40                bacterial cytoplasm --- Certainty=0.4582(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

- 45 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2892-

Example 2968

A DNA sequence (GASx2139) was identified in *S.pyogenes* <SEQ ID 8435> which encodes the amino acid sequence <SEQ ID 8436>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -5.89 Transmembrane 63 - 79 (54 - 80)

----- Final Results -----

bacterial membrane --- Certainty=0.3357(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2969

A DNA sequence (GASx2141R) was identified in *S.pyogenes* <SEQ ID 8437> which encodes the amino acid sequence <SEQ ID 8438>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4663(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2970

A DNA sequence (GASx2142) was identified in *S.pyogenes* <SEQ ID 8439> which encodes the amino acid sequence <SEQ ID 8440>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -10.08 Transmembrane 143 - 159 (135 - 165)

INTEGRAL Likelihood = -7.64 Transmembrane 53 - 69 (49 - 79)

INTEGRAL Likelihood = -7.17 Transmembrane 252 - 268 (248 - 275)

INTEGRAL Likelihood = -6.74 Transmembrane 186 - 202 (183 - 208)

INTEGRAL Likelihood = -5.63 Transmembrane 220 - 236 (218 - 240)

INTEGRAL Likelihood = -5.26 Transmembrane 116 - 132 (115 - 136)

INTEGRAL Likelihood = -2.02 Transmembrane 85 - 101 (85 - 101)

INTEGRAL Likelihood = -0.64 Transmembrane 165 - 181 (165 - 181)

----- Final Results -----

bacterial membrane --- Certainty=0.5034(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2893-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

5  >GP:AAD35257 GB:AE001701 conserved hypothetical protein [Thermotoga maritima]
    Identities = 81/275 (29%), Positives = 137/275 (49%), Gaps = 29/275 (10%)

    Query: 9  FKGMIIALGFILPGVSGGVLAAILGIYERMISFLAHMRDNFIENVLFFLPVGIG--GIL 65
              F G+++ + ++PGVSGG +A ++G+YE++I + ++ +PVG G G+
10  Sbjct: 7  FSGVLMGIANVVPVSGGTIAVLMGVYEKLIESVNSFFHCNSRSLKVLIPVGAGVLVGVF 66

    Query: 66  GIALFSFPVEFLKHYQSVLWGFAGAIVGTIPSLIKESTKQSQDKADWLWLVLTFVIS 125
              GIA F +E L Y V + F G I I S +K TK+ K + + FV+
15  Sbjct: 67  GIARF---LEIFLSKYPVPVTHFFFLGLI---IVSFVK--TKEYFSIKP---VNIFFVLL 114

    Query: 126 GLGLYFLNDLIG--TLPANFLTIFILAGALIALGVLVPGLSPLNLLLILGLYGPMLIGFKS 183
              G+ L F+ G T + +L G + A ++VPG+S S +LLI G+Y +L
20  Sbjct: 115 GMFLIFMLHFSGETTAKESMFLVLGGFVAATAMVVPGISGLILLIFGVYDHLVLYLVSH 174

    Query: 184 LDLLGTFLPLAIGGVLAAILAFSKSMDYALQHHHSKVYHFIIGIVLSSTLLILIPNSSSPE 243
              L ++G L +IG V IL K M++ L+ + Y FI G++L+S L ++P +
25  Sbjct: 175 L-IIGELLIFSIGVVAGILVSVKIMNFFLLKRFREETYSFIGGMILAS-LYEVLPPKMNTN 232

    Query: 244 SISYSHAGILTWLMAFVLFALGIWLGLWMSQLEEK 278
              + L + + L + LG ++ +E+K
30  Sbjct: 233 VV-----LPSVLSLVLSLTGLFFLLYIEKK 257

```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 2971

A DNA sequence (GASx2143R) was identified in *S.pyogenes* <SEQ ID 8441> which encodes the amino acid sequence <SEQ ID 8442>. Analysis of this protein sequence reveals the following:

Possible site: 20

35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

    bacterial cytoplasm --- Certainty=0.3964(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
40  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

45  >GP:BAB05000 GB:AP001511 unknown conserved protein in others
    [Bacillus halodurans]
    Identities = 28/78 (35%), Positives = 37/78 (46%)

    Query: 44  EVDKVFIVPLRQLLEFTDPVYRLEVTPIETTDFFPDRIRNGKYYQFSQEYRSIPFYENLE 103
              EVD VF VP+ + P YR+ V FP +RI N YQ S + FY
50  Sbjct: 127 EVDHVFTVPIDHFI SHPPEQYRINVHFEPGAGFPIERIANQSAYQKSTRQITESFYYYQS 186

    Query: 104 ETIWGMTAQFTKCLTDIL 121
              IWG+TA+ + + IL
55  Sbjct: 187 YVIWGLTAKILRHVITIL 204

```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2894-

Example 2972

A DNA sequence (GASx2144R) was identified in *S.pyogenes* <SEQ ID 8443> which encodes the amino acid sequence <SEQ ID 8444>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 17
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.4761(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

15 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2973

A DNA sequence (GASx2145) was identified in *S.pyogenes* <SEQ ID 8445> which encodes the amino acid sequence <SEQ ID 8446>. Analysis of this protein sequence reveals the following:

```

20   Possible site: 25

   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -4.09    Transmembrane    2 - 18 ( 1 - 19)

25   ----- Final Results -----
      bacterial membrane --- Certainty=0.2635(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

30 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAA49519 GB:X69895 X [Bacillus sphaericus]
Identities = 40/97 (41%), Positives = 57/97 (58%), Gaps = 5/97 (5%)

35   Query: 10  IEFLILAIVEKND SYGYDISQTIKLVAN----IKESTLYPILKKLEKAGFLTTYSQE-HQ 64
      ++ +IL ++ + D YGY+ISQ I   N   IKE+TLY + ++LEK   + Y   +
   Sbjct: 11  LDSIILRLILEKDRYGYEISQEISNRTNNSFQIKEATLYAVFQRLEKKEVIEAYYGDVSD 70

   Query: 65  GRKRKYAVTSSGRAQLIFLKKQSYKFALDGIIEG 101
      G KRKYY +TS G+A L   L KEW   K   +D +EG
40   Sbjct: 71  GGKRKYRITSLGKAYLSELVKQWAEVKEIIDLFMEG 107

```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2974

A DNA sequence (GASx2146) was identified in *S.pyogenes* <SEQ ID 8447> which encodes the amino acid sequence <SEQ ID 8448>. Analysis of this protein sequence reveals the following:

```

   Possible site: 56

50   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -14.75    Transmembrane    97 - 113 ( 77 - 143)

```

-2895-

```

INTEGRAL    Likelihood = -6.85    Transmembrane  116 - 132 ( 114 - 143)
INTEGRAL    Likelihood = -5.68    Transmembrane  156 - 172 ( 149 - 175)
INTEGRAL    Likelihood = -5.47    Transmembrane   79 -  95 (  77 -  96)

```

```

5  ----- Final Results -----
      bacterial membrane --- Certainty=0.6901(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

10 No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2975

15 A DNA sequence (GASx2147) was identified in *S.pyogenes* <SEQ ID 8449> which encodes the amino acid sequence <SEQ ID 8450>. Analysis of this protein sequence reveals the following:

Possible site: 31

```

20 >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -7.11    Transmembrane   8 -  24 (   6 -  30)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.3845(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
25      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```

30 >GP:AAF04457 GB:AF078161 lacunin [Manduca sexta]
      Identities = 68/310 (21%), Positives = 117/310 (36%), Gaps = 12/310 (3%)

Query: 55  DIDSSASTITVETGPFVQRPVTYYTHPKLIDPIVTTVTGKTLSSLQTPKDVVITGGIEIL 114
      DI+ + + + E+      T++ T   +   TT T T +S T + I   +
35 Sbjct: 1004 DIEGTTASGSTESTFTDETITMSKVTEESSVAEEETTKTITEEVSGTSESASINSDKITM 1063

Query: 115  GFTLNNRQEKNYRSIT--ITVPEKTSLNEVKASNPHTTSLNLT--VQDMQFDGNLTLL 170
      ++ +      IT +TV E+TS      TT+S ++   +      T
Sbjct: 1064 TTLS EDTGKTSVSEETITTEMTVTEETSETSPTEGTSKDTMTSTVSEETESSSVTEETTE 1123

40 Query: 171  HTKVKKATITGMLEATKSQLTNLELKADYSFSLNLTDSSE-NGTISLNGQLTKDITLK 229
      T V+ AT   E T S T +   ++ S   +++ E T +   T T+ K
Sbjct: 1124 TTVENATDISSTEVTASDKTTMTTMSSESEKTEEATTEITVKEVTESSSTETATSDK 1183

45 Query: 230  AVNIQSLHPGGIE-AERTTLENVFTVSKSKEEEEENDYDNDIAIFTAHALTLKGTNTITG 288
      ++ S G   AE +T E VT T   + EE      T+ +T+K T T
Sbjct: 1184 TISTLSEETGKTSVAEESITTEKVTETTVMPEETGK-----TITSEEITIKTTVTEEP 1237

Query: 289  GDIDVDITLTKAKAIAYRARTENGKVSLSQLTPAKIGKESTSDVISYVAENKAATGNLT 348
      D+   +T K   A E GK S+ + T   E++++ S A   T T
50 Sbjct: 1238 TDVGSSEAITSDKTTVSTASEETGKYSVSEETVKTVAEASTEPSSTEAITSDKTKMST 1297

Query: 349  VNLNKGDIIT 358
      ++ G ++
Sbjct: 1298 ISEETGKTSV 1307
55

```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2976

A DNA sequence (GASx2148R) was identified in *S.pyogenes* <SEQ ID 8451> which encodes the amino acid sequence <SEQ ID 8452>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2977

A DNA sequence (GASx2160) was identified in *S.pyogenes* <SEQ ID 8453> which encodes the amino acid sequence <SEQ ID 8454>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1630 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2978

A DNA sequence (GASx2170R) was identified in *S.pyogenes* <SEQ ID 8455> which encodes the amino acid sequence <SEQ ID 8456>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood =-13.32 Transmembrane 181 - 197 (175 - 203)

----- Final Results -----

bacterial membrane --- Certainty=0.6328 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2979

A DNA sequence (GASx2174) was identified in *S.pyogenes* <SEQ ID 8457> which encodes the amino acid sequence <SEQ ID 8458>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 28

   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -2.39    Transmembrane    3 - 19 ( 3 - 19)

   ----- Final Results -----
10      bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

15 Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2980

A DNA sequence (GASx2181R) was identified in *S.pyogenes* <SEQ ID 8459> which encodes the amino acid sequence <SEQ ID 8460>. Analysis of this protein sequence reveals the following:

```

20   Possible site: 24

   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
25      bacterial cytoplasm --- Certainty=0.3751(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

30 The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2981

35 A DNA sequence (GASx2185R) was identified in *S.pyogenes* <SEQ ID 8461> which encodes the amino acid sequence <SEQ ID 8462>. Analysis of this protein sequence reveals the following:

```

   Possible site: 26

   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.90    Transmembrane    18 - 34 ( 18 - 34)
40

   ----- Final Results -----
      bacterial membrane --- Certainty=0.1362(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
45      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

-2898-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2982

A DNA sequence (GASx2186R) was identified in *S.pyogenes* <SEQ ID 8463> which encodes the amino acid sequence <SEQ ID 8464>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4803 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA78948 GB:Z17279 transposase [*Streptococcus salivarius*]
Identities = 48/77 (62%), Positives = 55/77 (71%), Gaps = 1/77 (1%)

Query: 1 VSMKPIDLSKMSVIRKRSKKVMKTNKKTGKSIEERPEYINDRSEFGHWEIDLALGKKT 60
+ +K IDL + V IRK+ K T KK LGKSIEERPE IN+RS FG WEID LG KT
Sbjct: 150 LEIKVIDLPRAVRIRKKFTKRPST-KKHLGKSIEERPEINNRSRFGDWEIDSVLGGKTI 208

Query: 61 SEAVMLTLVERQTRYAL 77
E +LTLVERQTRYA+
Sbjct: 209 GEPSILTLVERQTRYAV 225

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2983

A DNA sequence (GASx2187R) was identified in *S.pyogenes* <SEQ ID 8465> which encodes the amino acid sequence <SEQ ID 8466>. Analysis of this protein sequence reveals the following:

Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3287 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA78948 GB:Z17279 transposase [*Streptococcus salivarius*]
Identities = 48/87 (55%), Positives = 57/87 (65%)

Query: 1 MNMSNINSTRKSSYSHLSATERGEIAAYLKMGGKPVETIARLLGSHRSTICREIKRGSVDQ 60
MNMS ST SY HLS ERGEI AYL +G KP EIAR LG +RSTI REI RGS+ Q
Sbjct: 1 MNMSTNYSTTNQSYKHLSEAERGEIEAYLSVGLKPAEIIARRLGRNRSTITREINRGSITQ 60

Query: 61 VKDKNGKQTFNFAYFADSRQRVYETNR 87
VK NG++ ++ Y+AD+ Y R
Sbjct: 61 VKKVNQKVYYQHYADAAHNRYRHAR 87

-2899-

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2984

A DNA sequence <SEQ ID 9013> was identified in *S.agalactiae* which encodes amino acid sequence
5 <SEQ ID 9014>. Analysis of the amino acid sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
McG: Discrim Score:      10.50
GvH: Signal Score (-7.5): -5.2
    Possible site: 40
10 >>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 4 value: -12.26 threshold: 0.0
    INTEGRAL    Likelihood = -12.26   Transmembrane  98 - 114 ( 94 - 116)
    INTEGRAL    Likelihood = -8.17    Transmembrane   5 - 21 ( 1 - 27)
    INTEGRAL    Likelihood = -6.95    Transmembrane  62 - 78 ( 57 - 80)
15    INTEGRAL    Likelihood = -5.84    Transmembrane  37 - 53 ( 30 - 55)
    PERIPHERAL  Likelihood = 17.35      81
    modified ALOM score: 2.95

*** Reasoning Step: 3
20 ----- Final Results -----
        bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01345(292 - 636 of 951)
PIR|G64646|G64646(56 - 168 of 205) hypothetical protein HP1015 - Helicobacter pylori
(strain 26695)
30 %Match = 4.4
    %Identity = 30.6 %Similarity = 54.1
    Matches = 34 Mismatches = 46 Conservative Sub.s = 26

87      117      147      177      207      237      267      297
35 LSGMGATFVPQTLIHRYLDKECNVYHFHKNKLFSEYIMYIKKDVELSGIALLLYKAFLTK*FR*FY*KSVYFLPKSV*NR
                                     |
                                     RYFLQNIHIIHQNKELQFIKKCLLGYFFAPLCGAILLVLFIVSSGAKSFQISNLFNN
                                     10      20      30      40      50

40 327      357      381      411      441      471      501
    PMYKIIASLFLVLPIFISQVL--VKIFKLKKNFIMFPDVAFPFIFVFLIPLISSLLKQNLPPYYLILISLLAIGITI--
    : | :: ||| : : : :: | : : | : || : :: | | : | : || : || :
    QLAYIVLLSLFLCALGFIAIGAIGFYRLSKITRHLSPFENFAFSFLAVILCAILSYLV-----PNASNALSLIGNGVISIFY
    70      80      90      100     110     120     130

45 549      579      606      636      666      696      726      756
    --KLLRTKTLFSYKRFLKLFWRSGF-ILTFLEYLGLLVIIIFIKVQ*KELDKLNCTPKVRQKI*RLGCFSDI*KL*R*TRN
    || | :|:: :|| : ||| :| | | || | :
50 LHKLYRELSLYTQERF----FLSGFRLLLFSLFMLALLGILVQALVVIIFLTAVVLMCVALGFLARAFLNFSQVFLKA
    150     160     170     180     190     200
```

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2985

55 A DNA sequence <SEQ ID 9015> was identified in *S.agalactiae* which encodes amino acid sequence
<SEQ ID 9016>. Analysis of the amino acid sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
```

-2900-

```
McG: Discrim Score:      13.20
GvH: Signal Score (-7.5): -2.08
    Possible site: 34
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value: 10.45 threshold: 0.0
    PERIPHERAL Likelihood = 10.45      36
    modified ALOM score:   -2.59
```

*** Reasoning Step: 3

----- Final Results -----

```

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

43.9/72.0% over 56aa

Streptococcus

pneumoniae

EGAD|7626| epua protein Insert characterized
SP|Q03159|EPUA_STRPN EPUA PROTEIN. Insert characterized
GP|47373|emb|CAA38133.1|X54225 7 kDa protein Insert characterized
PIR|S10640|S10640 epuA protein - Insert characterized

ORF01809(331 - 501 of 801)

EGAD|7626|7426(8 - 64 of 64) epua protein {Streptococcus pneumoniae}SP|Q03159|EPUA_STRPN
EPUA PROTEIN.GP|47373|emb|CAA38133.1||X54225 7 kDa protein {Streptococcus
pneumoniae}PIR|S10640|S10640 epuA protein - Streptococcus pneumoniae
%Match = 10.0

%Identity = 43.9 %Similarity = 71.9

Matches = 25 Mismatches = 16 Conservative Sub.s = 16

171 201 231 261 291 321 351 381
RSCLLTYELVQL*SWQEWLRK GKQ*L AN*PI*TVVIINS MN*KRLVLILNV*VRNMASSG WKYVLKQIGLIVLVILLA
:
MKMNKKSSYVVVKRLLLVIIVLILG

10 20

411 441 471 501 531 561 591 621
 LLFLAVGLMLGYSVFGDEGHAYSILSLDKWQNIIGKFLGK*KEPL*VI*CL*WFPLRVNFSSRIIQ*QKNK*QLRL*L
 | | : || : | | : : | : : || | || : | | |
 TLALGIGLMVGYGILGKGQDPWAILSPAKWQELIHKFTGN
 40 50 60

A related DNA sequence <SEQ ID 10507> was identified in GBS which encodes amino acid sequence <SEQ ID 10508>.

SEQ ID 9016 (GBS168) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 33 (lane 9; MW 7.6kDa) and in Figure 34 (lane 5; MW 7.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 42 (lane 2; MW 32.6kDa).

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vacc

Example 2986

A DNA sequence <SEQ ID 9017> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9018>. Analysis of the amino acid sequence reveals the following:

Lipop: Possible site: -1 Crend: 8
McG: Discrim Score: -2.85

-2901-

GvH: Signal Score (-7.5): -5.7
 Possible site: 21
 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 0 value: 5.25 threshold: 0.0
 PERIPHERAL Likelihood = 5.25 103
 modified ALOM score: -1.55

*** Reasoning Step: 3

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1210(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

56.1/72.0% over 131aa

Escherichia coli

EGAD|40237| arsenate reductase Insert characterized

SP|P52147|ARC2_ECOLI ARSENATE REDUCTASE (ARSENICAL PUMP MODIFIER). Edit characterized

GP|1061418|gb|AAB09628.1||U38947 ArsC {Plasmid R46} Insert characterized

ORF00095(304 - 699 of 1008)

EGAD|40237|42398(1 - 132 of 141) arsenate reductase {Escherichia coli} SP|P52147|ARC2_ECOLI
 ARSENATE REDUCTASE (ARSENICAL PUMP MODIFIER). GP|1061418|gb|AAB09628.1||U38947 ArsC
 {Plasmid R46}

%Match = 22.0

%Identity = 56.1 %Similarity = 72.0

Matches = 74 Mismatches = 37 Conservative Sub.s = 21

129	159	189	219	249	279	309	339
RIHSSL*PIFHRKRPYPSPRAFRFYFSNSCG*LWC*YCDWRELLAGLGINFYFLKTLVALKIERKMMKIRIYHNPNC							
						MSNITIYHNPHC	
							10
369	399	429	459	489	519	549	579
GTSRNVLAIRHCGIEPEITYYLKTPPSRMELVELLEMLKSARELLRTDVPAYEKFNLESSSVTDEEMIDAMIQDPILI							
	:	:	:	: :	:	:	
GTSRNTLEMIRNSGIEPTVILYLETPPSRDELLKLIADMGISVRALLRKKNVEPYEELGLAEDKFTDDQLIDFMLQHPILI							
	30	40	50	60	70	80	90
609	639	669	699	729	759	789	819
NRPIVVTSGAKLCRCPCEALTILPVKMEKDFVKEDGQIIQSL*HIV**IMV*EVSK*HY*KKLMRLETFCQKASQHQN							
		:	:	:	:	:	:
NRPIVVTPLGTLKCRPSEVVDILPDAQKAFTKEDGEKVVDSDSGKRLK							
	110	120	130	140			

SEQ ID 9018 (GBS45) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 8 (lane 4; MW 18.6kDa).

The GBS45-His fusion product was purified (Figure 97A; see also Figure 191, lane 5) and used to immunise mice (lane 1 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 97B), FACS (Figure 97C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2987

A DNA sequence <SEQ ID 9019> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9020>. Analysis of the amino acid sequence reveals the following:

-2902-

```

Lipop: Possible site: -1   Crend: 3
McG: Discrim Score:      6.84
GvH: Signal Score (-7.5): 2.98
    Possible site: 25
5  >>> Seems to have a cleavable N-term signal seq.
    ALOM program   count: 0 value: 13.69 threshold:  0.0
        PERIPHERAL Likelihood = 13.69      77
    modified ALOM score: -3.24

10  *** Reasoning Step: 3

    ----- Final Results -----
        bacterial outside --- Certainty=0.3000(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
15  bacterial cytoplasm --- Certainty=0.0000(Not Clear)

```

A DNA sequence <SEQ ID 10337> was identified in GBS which encodes amino acid sequence <SEQ ID 10338>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

SEQ ID 9020 (GBS55) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 7; MW 11.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 33 (lane 5; MW 36.3kDa).

GBS55-GST was purified as shown in Figure 197, lane 5.

GBS671 was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 161 (lane 2-4; MW 12kDa) and in Figure 188 (lane 2; MW 12kDa). Purified protein is shown in Figure 242, lane 3.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2988

A DNA sequence <SEQ ID 9021> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9022>. Analysis of the amino acid sequence reveals the following:

```

Lipop Possible site: -1   Crend: 3
McG: Discrim Score:      -14.35
GvH: Signal Score (-7.5): -2.12
    Possible site: 44
35  >>> Seems to have no N-terminal signal sequence
    ALOM program   count: 4 value: -13.90 threshold:  0.0
        INTEGRAL   Likelihood = -13.90   Transmembrane 101 - 117 ( 92 - 126)
        INTEGRAL   Likelihood = -7.64    Transmembrane 130 - 146 ( 125 - 148)
        INTEGRAL   Likelihood = -6.64    Transmembrane 24 - 40 ( 20 - 45)
40  INTEGRAL   Likelihood = -2.44    Transmembrane 55 - 71 ( 55 - 75)
        PERIPHERAL Likelihood = 17.40      2
    modified ALOM score:  3.28

    *** Reasoning Step: 3

45  ----- Final Results -----
        bacterial membrane --- Certainty=0.6562(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

SEQ ID 9022 (GBS215) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 175 (lane 10; MW 45kDa).

-2903-

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2989

A DNA sequence <SEQ ID 9023> was identified in *S.agalactiae* which encodes amino acid sequence
5 <SEQ ID 9024>. Analysis of the amino acid sequence reveals the following:

```

Lipop Possible site: -1   Crend: 0
McG: Discrim Score:      11.66
GvH: Signal Score (-7.5): -5.3
    Possible site: 61
10 >>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 2 value: -14.12 threshold: 0.0
    INTEGRAL    Likelihood = -14.12   Transmembrane  13 - 29 ( 5 - 35)
    INTEGRAL    Likelihood = -8.17    Transmembrane  44 - 60 ( 39 - 65)
    PERIPHERAL  Likelihood = 39.00     29
15 modified ALOM score: 3.32

```

*** Reasoning Step: 3

----- Final Results -----

```

20     bacterial membrane --- Certainty=0.6647(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

SEQ ID 9024 (GBS217) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell
25 extract is shown in Figure 85 (lane 2; MW 36.1kDa) and in Figure 156 (lane 1 & 3; MW 36kDa).

GBS217-GST was purified as shown in Figure 224, lane 5-6.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2990

30 A DNA sequence <SEQ ID 9025> was identified in *S.agalactiae* which encodes amino acid sequence
<SEQ ID 9026>. Analysis of the amino acid sequence reveals the following:

```

Lipop Possible site: -1   Crend: 10
McG: Discrim Score:      8.20
GvH: Signal Score (-7.5): -3.7
35     Possible site: 33
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 4 value: -9.98 threshold: 0.0
    INTEGRAL    Likelihood = -9.98   Transmembrane  22 - 38 ( 12 - 43)
    INTEGRAL    Likelihood = -7.80   Transmembrane  61 - 77 ( 56 - 85)
40     INTEGRAL    Likelihood = -5.20   Transmembrane  121 - 137 ( 117 - 148)
    INTEGRAL    Likelihood = -2.97   Transmembrane  99 - 115 ( 98 - 119)
    PERIPHERAL  Likelihood = 10.77     5
    modified ALOM score: 2.50

```

*** Reasoning Step: 3

----- Final Results -----

```

45     bacterial membrane --- Certainty=0.4991(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence <SEQ ID 10701> was identified in GBS which encodes amino acid sequence
<SEQ ID 10702>.

-2904-

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2991

A DNA sequence <SEQ ID 9027> was identified in *S.agalactiae* which encodes amino acid sequence
5 <SEQ ID 9028>. Analysis of the amino acid sequence reveals the following:

```
Lipop Possible site: -1   Crend: 7
McG: Discrim Score:      10.61
GvH: Signal Score (-7.5): -4.21
Possible site: 51
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 3 value: -10.99 threshold: 0.0
  INTEGRAL     Likelihood = -10.99   Transmembrane  38 - 54 ( 33 - 61)
  INTEGRAL     Likelihood = -8.01    Transmembrane   5 - 21 ( 1 - 26)
  INTEGRAL     Likelihood = -7.01    Transmembrane  65 - 81 ( 60 - 87)
15 PERIPHERAL   Likelihood = 13.85      99
modified ALOM score: 2.70
```

*** Reasoning Step: 3

----- Final Results -----

```
bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful
25 antigens for vaccines and/or diagnostics.

Example 2992

A DNA sequence <SEQ ID 9029> was identified in *S.agalactiae* which encodes amino acid sequence
30 <SEQ ID 9030>. Analysis of the amino acid sequence reveals the following:

```
Lipop Possible site: -1   Crend: 10
McG: Discrim Score:      -21.39
GvH: Signal Score (-7.5): -1.85
Possible site: 57
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value: -8.44 threshold: 0.0
35 INTEGRAL     Likelihood = -8.44    Transmembrane  38 - 54 ( 36 - 59)
  PERIPHERAL   Likelihood = 19.10      18
modified ALOM score: 2.19
```

*** Reasoning Step: 3

----- Final Results -----

```
bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful
45 antigens for vaccines and/or diagnostics.

Example 2993

A DNA sequence <SEQ ID 9031> was identified in *S.agalactiae* which encodes amino acid sequence
50 <SEQ ID 9032>. Analysis of the amino acid sequence reveals the following:

```
Lipop Possible site: -1   Crend: 5
McG: Discrim Score:      12.87
GvH: Signal Score (-7.5): -3.57
Possible site: 41
```

-2905-

```

>>> Seems to have an uncleavable N-term signal seq
ALOM program    count: 4 value: -10.30 threshold: 0.0
  INTEGRAL      Likelihood = -10.30  Transmembrane  69 - 85 ( 63 - 98)
  INTEGRAL      Likelihood = -8.65   Transmembrane   4 - 20 (  1 - 29)
  INTEGRAL      Likelihood = -2.07   Transmembrane  96 - 112 ( 95 - 118)
  PERIPHERAL    Likelihood = 9.71    113
modified ALOM score: 2.56

```

```

*** Reasoning Step: 3

```

```

----- Final Results -----

```

```

      bacterial membrane --- Certainty=0.5118(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

20.1/50.5% over 114aa

Streptococcus pneumoniae

GP|9798572| BlpX protein Insert characterized

ORF02100(316 - 660 of 999)

GP|9798572|emb|CAC03527.1||AJ276410(9 - 123 of 132) BlpX protein {Streptococcus pneumoniae}

%Match = 5.0

%Identity = 20.0 %Similarity = 50.4

Matches = 23 Mismatches = 57 Conservative Sub.s = 35

```

90      120      150      180      210      240      270      300
LMSLF*DPQVSGEELDKFTVRLDSHRKSNSRG*NQLVIIRLYSQIN*REPNNLVGPFLNKGEHMTQDYICYL*SRGGED

```

MEV

```

330      360      390      420      450      480      510      540
MHNILRFLGIVIIISAVILFSIGSFYDLTLMKNILLICWSFLFDLLVFVFKQRQTTEVLTWYQVVKQFWLFIKCTILIPIL
:  |: :: : |: |: : |: |: : :: : |: :| |: :| : : : : : : : : : : : : : : : : : : : : : :
FNMKYRLFFVIFLSSVLDILLGTFLQISIVSIGWLVLYSGLFEAGVFLANKGVAVKIKEVDIRNRFKFIKFTLWFQIL
      20      30      40      50      60      70      80

```

```

570      600      630      660      690      720      750      780
VAFIIMKGCLTSISDILIIYFYLHLVVVYYTIGMILSLGRIISPEHSMFNKLRK*NELYLKVFVFNRADLTICCLPCLS*FF
:  :: : : : || || |: |: : : || : : : : : : : : : : : : : : : : : : : : : :
LLIFLIILKLYLGLDARLILFYGHIFVFNALMYLLSSSQVSLKKNKLSS
      100      110      120      130

```

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2994

A DNA sequence <SEQ ID 9033> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9034>. Analysis of the amino acid sequence reveals the following:

Lipop Possible site: -1 Crend: 9

McG: Discrim Score: 3.25

GvH: Signal Score (-7.5): -3.39

Possible site: 59

>>> Seems to have an uncleavable N-term signal seq

ALOM program count: 4 value: -6.64 threshold: 0.0

INTEGRAL Likelihood = -6.64 Transmembrane 46 - 62 (43 - 64)

INTEGRAL Likelihood = -5.15 Transmembrane 17 - 33 (15 - 34)

PERIPHERAL Likelihood = 11.03 100

modified ALOM score: 1.83

```

*** Reasoning Step: 3

```

```

----- Final Results -----

```

-2906-

```

bacterial membrane --- Certainty=0.3654(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

5 The protein has homology with the following sequences in the databases:

35.5/63.8% over 127aa

OMNI|NT01BS4455| wall teichoic acid glycosylation protein GtcA Insert characterized

```

10      ORF01715(343 - 750 of 1053)
      OMNI|NT01BS4455(58 - 185 of 187) wall teichoic acid glycosylation protein GtcA
      %Match = 8.0
      %Identity = 35.5  %Similarity = 63.7
      Matches = 44  Mismatches = 39  Conservative Sub.s = 35

15      210          240          270          300          330          360          390          411
      GN*ASRVV*NNLLSISQTKSKAKLMGDFLITLKHP*YKNMVKLKSLLKKSIQNEVSLYLLFGLLTSLLYLV---IRQGI
          :  :          :          :  : |  :          :  :          |  :  :  :  :  :  :  :  :  :  :  :
      PRRNHQTIICIGPASHLPQLFRRTLGIIFYFRQRAAREAKNFEKFFRKRGTSVKYREIIMYIIMGVFTTIVNIASFYILVEI

20          20          30          40          50          60          70          80

      441          471          501          531          549          579          609          639
      FNFSQDAPFSAIVANIIAILFAFFTNDRFVFKQTKIEQLQLR---QTFVIARLGTGLDLILAVIFVDQFPSIIGQFVQ
      |  |  :  :  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
25      MNVDYKA--ATVAAWILSVLFAYITNKLYVFFQ--KTHDLQSLKELTAFFSVRVLSLIGIDLGMMIILVGQF-----
          100          110          120          130          140          150

      669          690          720          750          780          810          840          870
      HNLNKKINTIESL---VSQILITLLNYILSKFVIFKDKKRQL*QELSFLIFLLWIFG*ET*YLHALIQFFLSQFLERWHSV
          |  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
30      -----NTNETLAKILDNAVIVVVVYVASKWLVFKKTKKEGV
          160          170          180

```

SEQ ID 9034 (GBS283) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 63 (lane 8; MW 67.6kDa).

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2995

A DNA sequence <SEQ ID 9035> was identified in *S.agalactiae* which encodes amino acid sequence
40 <SEQ ID 9036>. Analysis of the amino acid sequence reveals the following:

```

Lipop Possible site: -1    Crend: 2
SRCFLG: 0
McG: Length of UR:    22
      Peak Value of UR:    3.86
      Net Charge of CR: 2
McG: Discrim Score:    16.84
GvH: Signal Score (-7.5): -4.38
      Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program    count: 1 value: -12.37 threshold: 0.0
      INTEGRAL    Likelihood =-12.37    Transmembrane    7 - 23 ( 1 - 26)
      PERIPHERAL  Likelihood = 12.84    64
      modified ALOM score: 2.97
icm1 HYPID: 7    CFP: 0.595

*** Reasoning Step: 3

----- Final Results -----
bacterial membrane --- Certainty=0.5946(Affirmative) < succ

```


-2907-

```

bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

SEQ ID 9036 (GBS286) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 11; MW 16.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 59 (lane 2; MW 41.3kDa) and in Figure 63 (lane 9; MW 41.4kDa).

The GBS286-GST fusion product was purified (Figure 210, lane 9; Figure 225, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 274), which confirmed that the protein is immunoaccessible on GBS bacteria.

GBS668 was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 139 (lane 2-4; MW 43.5kDa) and in Figure 187 (lane 6; MW 43kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 139 (lane 6 & 7; MW 18.6kDa) and in Figure 179 (lane 12; MW 19kDa).

GBS668-GST was purified as shown in Figure 237 (lane 10). GBS668-His was purified as shown in Figure 231 (lanes 5 & 6).

GBS673 was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 161 (lane 8-10; MW 17kDa) and in Figure 188 (lane 4; MW 17kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 162 (lane 8; MW 41.5kDa) and in Figure 239 (lane 7; MW 41kDa). Purified GBS673-His is shown in Figure 242, lane 5. Purified GBS673-GST is shown in Figure 246, lane 2.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2996

A DNA sequence <SEQ ID 9037> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9038>. Analysis of the amino acid sequence reveals the following:

```

Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:      -18.42
GvH: Signal Score (-7.5): -6.16
    Possible site: 57
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 2 value: -8.49 threshold: 0.0
    INTEGRAL    Likelihood = -8.49   Transmembrane  51 - 67 ( 44 - 95)
    INTEGRAL    Likelihood = -3.08   Transmembrane  70 - 86 ( 68 - 95)
    PERIPHERAL  Likelihood = 12.89    32
modified ALOM score: 2.20

*** Reasoning Step: 3

----- Final Results -----
    bacterial membrane --- Certainty=0.4397(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

SEQ ID 9038 (GBS386) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 70 (lane 2; MW 14kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 8; MW 39.5kDa).

-2908-

GBS386-GST was purified as shown in Figure 213, lane 8.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2997

- 5 A DNA sequence <SEQ ID 9039> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9040>. Analysis of the amino acid sequence reveals the following:

```

Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      -15.47
GvH: Signal Score (-7.5): -6.21
10   Possible site: 14
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 2 value: -3.61 threshold: 0.0
    INTEGRAL   Likelihood = -3.61   Transmembrane   94 - 110 ( 94 - 111)
    INTEGRAL   Likelihood = -1.70   Transmembrane   75 - 91 ( 75 - 91)
15   PERIPHERAL Likelihood = 5.94    139
    modified ALOM score: 1.22

*** Reasoning Step: 3

20   ----- Final Results -----
        bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
        bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25   The protein has homology with the following sequences in the databases:

    ORF01481(394 - 720 of 1065)
    GP|9657521|gb|AAF96047.1||AE004354(16 - 121 of 243) uridine phosphorylase {Vibrio cholerae}
    %Match = 5.3
    %Identity = 28.0 %Similarity = 48.6
30   Matches = 30 Mismatches = 54 Conservative Sub.s = 22

    150      180      210      240      270      300      330      360
    V*KH MV*AI*YGNLP*KW*IVPLSIFIFANLTLPFKFH*VKIEKIFLTR**NIVN*GLKEMLMINSFDNSRKAIINPED

35                                          MSIQ

    390      420      450      480      510      540      570      600
    INSPIKGFPKTVITCFARETFNRILEEPLPHREIARTSVANLEIPIYELEFKGQKIGFFNAYVGASACVAILEDIIVFGME
    | : |      |||      | : |      | | : : || : | | : : || : : | :
40   PHIHVAQVAPRVVCGEPNRRANRIASLLNNAE---LVAENREYRLFSGEFEEQPTVCSTGIGAPSMIIAVEELARSGAK
        20      30      40      50      60      70      80

    630      660      690      720      750      780      810      840
    SLIVFGTCGVLDSSIEETSIIIPRSAIRDEGTSFHYSEASSEIAVNTNSIFLLCG*FRCRSMGSKI FRK*RGFRKER*NC
    : : | : | : | : | : | : | : | : | : | : | : | : | : | :
45   AIVRVGSAGAMQSEIGLGELILVEGAVRDEGGSKAYIGAAYPAYSSFELVVEMQRF LAEQSVPIHRGIVRSHDSFYTDEE
        100      110      120      130      140      150      160

```

- 50 SEQ ID 9040 (GBS388) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 70 (lane 3; MW 21kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 9; MW 45.6kDa).

The GBS388-GST fusion product was purified (Figure 213, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 311), which confirmed that the protein is immunoaccessible on GBS bacteria.

-2909-

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2998

A DNA sequence <SEQ ID 9041> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9042>. Analysis of the amino acid sequence reveals the following:

```

Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:    -11.81
GvH: Signal Score (-7.5): -7.49
    Possible site: 25
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value: -5.68 threshold: 0.0
    INTEGRAL   Likelihood = -5.68   Transmembrane   78 - 94 ( 77 - 95)
    PERIPHERAL Likelihood = 4.61    134
modified ALOM score: 1.64

*** Reasoning Step: 3

----- Final Results -----
    bacterial membrane --- Certainty=0.3272(Affirmative) < succ>
    bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

ORF01912(307 - 720 of 1056)
GP|3845252|gb|AAC71927.1||AE001412(81 - 242 of 244) hypothetical protein {Plasmodium
falciparum} PIR|D71608|D71608 hypothetical protein PFB0690w - malaria parasite (Plasmodium
falciparum)
%Match = 4.0
%Identity = 31.2 %Similarity = 53.5
Matches = 45 Mismatches = 58 Conservative Sub.s = 32

231      261      291      348      378      405
KKGRFLIDLVCNVMNFKNSKIA*NQCFDV**RVVNHLIN-LSKENIAKIDFDLNEALNA-NIRLKELVDELKISK----
||| : | :: : | : | : : || | | || |
KYNELQSLLSKEEEKYDFVKNELGDLQKQKDLLKWHLCNNIKKLSMKRSDYKFKTETKSKLESKLSKSLKDMNKIKHFEHD
60      70      80      90      100     110     120

450      480      501      531      558      588      618
-----ELDSKGWSKKDSRTIKILYDGLINK--HIVSLDRADYNII-QVIPFANVHVLFLIPERENSKNYRIY
||::| : | | | : :: ||| :: : : :: : | : || :| :| :| :|
TLEELVHKMEQELETIKMYIKND---IENIFNECINKKDEYLYKDITQERISVFKERKKRQNLQKLLIMKQENNKNYNIN
140     150     160     170     180     190     200

648      672      693      720      750      780      810      840
NYS DYEMELINE--DRQQFSKYET---VDL-DQLILVDIFNIDDISSYLT*DIENLDLGLLKLINYADNKS DRHILQT
|| | : || : : : || : || | | : :
YLK KYESNL MNEINSYKNYKDFETKIAMDLDHSLNDLYVT
220     230     240

```

A related DNA sequence <SEQ ID 10589> was identified in GBS which encodes amino acid sequence <SEQ ID 10590>.

SEQ ID 9042 (GBS408) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 6; MW 20.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 171 (lane 5; MW 45.3kDa).

GBS408-GST was purified as shown in Figure 218, lane 9.

-2910-

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 2999

A DNA sequence <SEQ ID 9043> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9044>. Analysis of the amino acid sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 9
McG: Discrim Score:      -9.62
GvH: Signal Score (-7.5): -4.84
    Possible site: 61
10 >>> Seems to have no N-terminal signal sequence
ALOM program    count: 2 value: -11.09 threshold: 0.0
    INTEGRAL      Likelihood ==-11.09    Transmembrane    45 - 61 ( 37 - 72)
    INTEGRAL      Likelihood = -8.60     Transmembrane    76 - 92 ( 70 - 97)
    PERIPHERAL    Likelihood = 11.62      95
15 modified ALOM score: 2.72

*** Reasoning Step: 3

----- Final Results -----
20          bacterial membrane --- Certainty=0.5437(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

25 ORF01977(442 - 627 of 948)
   EGAD|88220|96064(204 - 583 of 751) hypothetical 848 kDa protein f23f125 in chromosome iii
   {Caenorhabditis elegans} SP|P46501|YLK5_CAEEL HYPOTHETICAL 84.8 KDA PROTEIN F23F12.5 IN
   CHROMOSOME III. GP|529214|gb|AAA20607.1||U12965 F23F12.5 gene product {Caenorhabditis
   elegans}
30 %Match = 4.6
   %Identity = 35.9 %Similarity = 59.4
   Matches = 23 Mismatches = 24 Conservative Sub.s = 15

192      222      252      282      312      342      372      402
35 DFVSSFFIS*SQTNYNRISFLLKLAKHQLECLNNVAQGLSV**YSSMKDYINRILHFIKEHMTYHVNFIDDFLDIKWEKV

VTL SAYFPFTITVERYAMNKSEKYEKMPIILGPLFVLFIKLELKIKDKVTLFQVIVNFGVIFQIYKNETFSHGDAFAS
      120      130      140      150      160      170      180      190

40 432      462      492      522      552
   SNIHLRFWTTIIAYLVIFILSISTVILNLVLLFQGFLTONPIIYLLFFITLVCAFY-----
   |:|:|::: |:|:|:| | | :| |:| |:
   LYPPGTAEKVFTFYVVLFLINLLDVMFNVLVLLQMSFLNTRNFHWLCFFLWQFALFFCCQQIFSIFYNFSPLGSCDD~~~~
      200      210      220      230      240      250      260

45 600      627      657      687      717      747
   -----FAYKFITYTPTIVKNAL-QYIKKLKNV*NNKVICTLTLYQLCFRVFHTKITKKDSYLTII
   | : : | | : : | |
   AGNFYLSQFVSGAVTAFAKIFVFLDDTYVPSFDRRLHQYPPQIAMILCYCIMVLMILPESDCGSGQSRDLAIIINIIG
50      560      570      580      590      600      610      620

```

SEQ ID 9044 (GBS411) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 78 (lane 2; MW 16kDa).

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful
55 antigens for vaccines and/or diagnostics.

-2911-

Example 3000

A DNA sequence <SEQ ID 9045> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9046>. Analysis of the amino acid sequence reveals the following:

```

5   Lipop: Possible site: -1   Crend: 6
   McG: Discrim Score:      -17.94
   GvH: Signal Score (-7.5): -4.63
       Possible site: 45
   >>> Seems to have no N-terminal signal sequence
   ALOM program   count: 1 value: -6.10 threshold: 0.0
10  INTEGRAL      Likelihood = -6.10   Transmembrane  31 - 47 ( 26 - 49)
   PERIPHERAL    Likelihood = 15.33     3
   modified ALOM score: 1.72

   *** Reasoning Step: 3
15  ----- Final Results -----
       bacterial membrane --- Certainty=0.3442(Affirmative) < succ>
       bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
20  bacterial cytoplasm  --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

25  ORF01982(313 - 501 of 801)
   GP|2444082|gb|AAC79518.1||U88974(93 - 156 of 156) ORF2 {Streptococcus thermophilus
   temperate bacteriophage 01205} PIR|T13290|T13290 hypothetical protein 2 - Streptococcus
   phage phi-01205
   %Match = 11.5
   %Identity = 48.4 %Similarity = 59.4
   Matches = 31 Mismatches = 25 Conservative Sub.s = 7

30  174      204      234      264      294      324      354      384
   DVDQNIESHKLFKRHFV*RAILPQSKRK*EN**LCVISEPR*KLKSKLGELKMGFFAQRCPYCQSTKVQFMNQDRKGFNG
                                     | :|||:| | |:| | | | :
   LLMFVGVALLFARLFWEIKHPMTKEQKEQLKIERAKAKEEFRKSRNEFKKAMAEARAVKCPYCKSTDVEFMVQQRKSFSI
35  50      60      70      80      90      100     110
   CVGCIGFLIAWPF-LLLGLVGKKGKNNWHCTNCGRTFKTK*KSPTLKFCPRRA*GKF*YSKNLLFGRGFYHTYFNRK*GI
   | ::      | : |||| | || | | | |
40  GKAAAGTIMTGGVGALAGFAGKKGKKEWHCKNCGAVFTTK
       130      140      150

```

SEQ ID 9046 (GBS412) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 171 (lane 6; MW 36kDa). Purified GBS412-GST is shown in Figure 218, lane 10-11.

45 Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3001

A DNA sequence <SEQ ID 9047> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9048>. Analysis of the amino acid sequence reveals the following:

```

50  Lipop: Possible site: -1   Crend: 0
   McG: Discrim Score:      3.67
   GvH: Signal Score (-7.5): -3.62
       Possible site: 41
   >>> Seems to have an uncleavable N-term signal seq
55  ALOM program   count: 5 value: -7.27 threshold: 0.0
   INTEGRAL      Likelihood = -7.27   Transmembrane  48 - 64 ( 32 - 68)
   INTEGRAL      Likelihood = -6.26   Transmembrane  87 - 103 ( 85 - 105)

```

-2912-

```

INTEGRAL    Likelihood = -6.21    Transmembrane    29 - 45 ( 26 - 46)
INTEGRAL    Likelihood = -3.29    Transmembrane    110 - 126 ( 109 - 130)
INTEGRAL    Likelihood = -2.87    Transmembrane     2 - 18 ( 1 - 18)
PERIPHERAL  Likelihood = 4.24      66
modified ALOM score: 1.95

```

*** Reasoning Step: 3

----- Final Results -----

```

bacterial membrane --- Certainty=0.3909(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

ORF01286(304 - 672 of 993)
GP|8272442|dbj|BAA96471.1|AB036428(90 - 212 of 218) type IV prepilin peptidase homologue
{Streptococcus mutans}
%Match = 16.8
%Identity = 46.3 %Similarity = 72.4
Matches = 57 Mismatches = 34 Conservative Sub.s = 32

```

```

102      132      162      192      222      252      282      312
*RRLPIT*IPNFFKRFCTSNKTFIYEF*QKTIQFSRKSATAC*LSL*R*TDYL**KS*SLFYHFSNININYKKDFMIMS
                                                                    :::
LGSFFGLVVD RYPQKSIIFPRSHCNKCYNCLTMRDLIPFISRIINKNSCRFCGYPIPLRYSLVVELLCGLISTGFALDLLT
          30          40          50          60          70          80          90

```

```

342      372      402      432      462      492      522      552
TIYFISLCMSFILSYDDIKYQEIPIFLWILFTISTITLTPTKVSIVLCLFGILAEVVDINIGSGDFLYLATIGLSLPLH
|      |      :||  ||:  ||:  |||  ||      :  :  :  :|||:  :  :|||:  :|||:  |||  |:
TSQVCLLFMGVLLSLYDLQDSYPLTLWIGFTFLIMFIYPLNLISLILFLFGIFAALKNINIGSGDFFYLATLALSINLQ
          110         120         130         140         150         160         170

```

```

582      612      642      672      702      732      762      792
QMLFIIQIGAWLGIIYCLVMRKMKKTFIAFLPFLSIAYIIVTSYSLLF*SL*IIRKVIKLWVLVAFWIFRMTNCTTKINHG
|:::||||  :  |||:  |:  :  |:  ||:||||  :  ::|:
QIIWIIQIASLLGILYSLLFQKHKEPFAFVPFLFLGHLIIIFSHLI
          190         200         210

```

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3002

A DNA sequence <SEQ ID 9049> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9050>. Analysis of the amino acid sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 2
McG: Discrim Score: 10.43
GvH: Signal Score (-7.5): -4.39
    Possible site: 54
>>> Seems to have an uncleavable N-term signal seq
ALOM program    count: 4 value: -10.30 threshold: 0.0
INTEGRAL    Likelihood = -10.30    Transmembrane    62 - 78 ( 59 - 84)
INTEGRAL    Likelihood = -6.10    Transmembrane     4 - 20 ( 1 - 22)
INTEGRAL    Likelihood = -4.25    Transmembrane   128 - 144 ( 123 - 145)
INTEGRAL    Likelihood = -3.13    Transmembrane    88 - 104 ( 87 - 104)
PERIPHERAL  Likelihood = 2.01      109
modified ALOM score: 2.56

```

*** Reasoning Step: 3

----- Final Results -----

```

bacterial membrane --- Certainty=0.5118(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-2913-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

```

SP|Q48729|LSPA_LACLC(1 - 149 of 150) LIPOPROTEIN SIGNAL PEPTIDASE (EC 3.4.23.36)
5 (PROLIPOPROTEIN SIGNAL PEPTIDASE) (SIGNAL PEPTIDASE II) (SPASE II).
%Match = 16.3
%Identity = 40.7 %Similarity = 66.0
Matches = 61 Mismatches = 50 Conservative Sub.s = 38

10      180          210          240          270          300          330          360          390
EWYHCYSIRSSSR*PNDLYQTKRS*FISDGFKICCCYGRVF*GI*FIGEVMRKIIPITITILLIALDQLSKLWIVKHIEL
                                     |:|:: :| :: | ||: | |:| ::|
                                     MKKLLSLVIIVVGIIADQVFNKWVVANIQL
                                                10          20          30

15      420          450          480          510          540          570          600          630
NQIKEFIPNIVSLTYLRNYGAAFSILQNQQWLFTLTITIFVVGVAIYYLMKHINGSYWLLISLTLIISGGLGNFIDRLRLG
   |: |:::|||::| |||:| : |||:| ::| |:|:::| | | : ||||:| |||:: |:| |
GDTKKIWPDVLSTLYIKNDGAAWSSFSGQQWFFLVLTPIVLIVALWFLWK-KMGONWYFAGLTTLIAGALGNLLTRVRQG
               40          50          60          70          80          90          100

20      660          690          720          750          780          810          840          870
YVVDMVHLDFINF AIFNVADSYLTIGIICLMIALWKEESNGNHN*NSRS*AR*SFGS*F*TVA*SS**RN*KRDCVSKWT
:|||| :| |||:| |::| : | ||: ::
FVVDMEFXNRIYDFPIFNIAIDILLSVGFVVLFAILTDEKET
                120          130          140          150

```

There is also homology to SEQ ID 7750.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful
30 antigens for vacc

Example 3003

A DNA sequence <SEQ ID 9051> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9052>. Analysis of the amino acid sequence reveals the following:

```

35 Lipop: Possible site: -1      Crend: 7
    McG: Discrim Score:      13.24
    GvH: Signal Score (-7.5): -2.18
        Possible site: 19
    >>> Seems to have a cleavable N-term signal seq.
    ALOM program   count: 0 value: 2.01 threshold: 0.0
40     PERIPHERAL  Likelihood = 2.01      21
        modified ALOM score: -0.90

    *** Reasoning Step: 3

45 ----- Final Results -----
        bacterial outside --- Certainty=0.3000(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

50 The protein has no homology with any sequences in the databases.

SEQ ID 9052 (GBS138) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 2; MW 15kDa)

GBS672 was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 161 (lane 5-7; MW 15kDa) and in Figure 188 (lane 3; MW 15kDa). Purified protein is shown in Figure 242, lane 4.

-2914-

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3004

A DNA sequence <SEQ ID 9053> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9054>. Analysis of the amino acid sequence reveals the following:

```

Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      18.01
GvH: Signal Score (-7.5): -2.35
    Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value: 14.80 threshold: 0.0
    PERIPHERAL Likelihood = 14.80      51
    modified ALOM score: -3.46

*** Reasoning Step: 3

----- Final Results -----
        bacterial outside --- Certainty=0.3000(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

SEQ ID 9054 (GBS143) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 2; MW 33.5kDa).

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3005

A DNA sequence <SEQ ID 9055> was identified in *S.agalactiae* which encodes amino acid sequence <SEQ ID 9056>. Analysis of the amino acid sequence reveals the following:

```

Lipop Possible site: -1   Crend: 0
McG: Discrim Score:      7.43
GvH: Signal Score (-7.5): -6.25
    Possible site: 41
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value: -10.77 threshold: 0.0
    INTEGRAL    Likelihood = -10.77   Transmembrane    2 - 18 ( 1 - 20)
    PERIPHERAL Likelihood = 5.14      29
    modified ALOM score: 2.65

*** Reasoning Step: 3

----- Final Results -----
        bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

SEQ ID 9056 (GBS229) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 3; MW 35.9kDa).

GBS229-GST was purified as shown in Figure 206, lane 5.

Based on this analysis, it is predicted that this protein from *S.agalactiae*, and its epitopes, could be useful antigens for vaccines and/or diagnostics.

-2915-

Example 3006

A DNA sequence <SEQ ID 9183> was identified in GAS which encodes amino acid sequence <SEQ ID 9184>. Analysis of the amino acid sequence reveals the following:

Possible site: 29

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3007

A DNA sequence <SEQ ID 9185> was identified in GAS which encodes amino acid sequence <SEQ ID 9186>. Analysis of the amino acid sequence reveals the following:

Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3008

A DNA sequence <SEQ ID 9187> was identified in GAS which encodes amino acid sequence <SEQ ID 9188>. Analysis of the amino acid sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.70 Transmembrane 850 - 866 (850 - 866)
INTEGRAL Likelihood = -1.22 Transmembrane 15 - 31 (15 - 31)

----- Final Results -----

bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3009

A DNA sequence <SEQ ID 9189> was identified in GAS which encodes amino acid sequence <SEQ ID 9190>. Analysis of the amino acid sequence reveals the following:

LPXTG motif: 259-263

Possible site: 13

-2916-

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.93 Transmembrane 270 - 286 (268 - 288)

----- Final Results -----

5 bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for
 10 vaccines and/or diagnostics.

Example 3010

A DNA sequence <SEQ ID 9191> was identified in GAS which encodes amino acid sequence <SEQ ID 9192>. Analysis of the amino acid sequence reveals the following:

Possible site: 21

15 >>> May be a lipoprotein

----- Final Results -----

20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for
 vaccines and/or diagnostics.

25 Example 3011

A DNA sequence <SEQ ID 9193> was identified in GAS which encodes amino acid sequence <SEQ ID 9194>. Analysis of the amino acid sequence reveals the following:

Possible site: 29

30 >>> May be a lipoprotein

----- Final Results -----

35 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for
 vaccines and/or diagnostics.

Example 3012

40 A DNA sequence <SEQ ID 9195> was identified in GAS which encodes amino acid sequence <SEQ ID 9196>. Analysis of the amino acid sequence reveals the following:

Possible site: 34

45 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

50 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for
 vaccines and/or diagnostics.

-2917-

Example 3013

A DNA sequence <SEQ ID 9197> was identified in GAS which encodes amino acid sequence <SEQ ID 9198>. Analysis of the amino acid sequence reveals the following:

```

Possible site: 13
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -3.50    Transmembrane  346 - 362 ( 343 - 366)
    INTEGRAL    Likelihood = -2.97    Transmembrane  177 - 193 ( 176 - 195)

----- Final Results -----
10      bacterial membrane --- Certainty=0.2402(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3014

A DNA sequence <SEQ ID 9199> was identified in GAS which encodes amino acid sequence <SEQ ID 9200>. Analysis of the amino acid sequence reveals the following:

```

Possible site: 19
20  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -1.33    Transmembrane  24 - 40 ( 24 - 40)

----- Final Results -----
25      bacterial membrane --- Certainty=0.1532(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3015

A DNA sequence <SEQ ID 9201> was identified in GAS which encodes amino acid sequence <SEQ ID 9202>. Analysis of the amino acid sequence reveals the following:

```

Possible site: 33
35  >>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -6.00    Transmembrane  194 - 210 ( 192 - 214)

----- Final Results -----
40      bacterial membrane --- Certainty=0.3399(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 183-187

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3016

A DNA sequence <SEQ ID 9203> was identified in GAS which encodes amino acid sequence <SEQ ID 9204>. Analysis of the amino acid sequence reveals the following:

```

Possible site: 32
50  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood =-11.25    Transmembrane  9 - 25 ( 4 - 28)

```

-2918-

----- Final Results -----

5 bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3017

10 A DNA sequence <SEQ ID 9205> was identified in GAS which encodes amino acid sequence <SEQ ID 9206>. Analysis of the amino acid sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

15 INTEGRAL Likelihood = -3.03 Transmembrane 462 - 478 (460 - 479)
 INTEGRAL Likelihood = -0.90 Transmembrane 18 - 34 (18 - 34)

----- Final Results -----

20 bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 450-454

25 Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3018

A DNA sequence <SEQ ID 9207> was identified in GAS which encodes amino acid sequence <SEQ ID 9208>. Analysis of the amino acid sequence reveals the following:

Possible site: 26

30 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -2.60 Transmembrane 15 - 31 (12 - 32)

----- Final Results -----

35 bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3019

A DNA sequence <SEQ ID 9209> was identified in GAS which encodes amino acid sequence <SEQ ID 9210>. Analysis of the amino acid sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

45 INTEGRAL Likelihood = -1.38 Transmembrane 16 - 32 (16 - 32)

----- Final Results -----

50 bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2919-

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3020

A DNA sequence <SEQ ID 9211> was identified in GAS which encodes amino acid sequence <SEQ ID 9212>. Analysis of the amino acid sequence reveals the following:

```

Possible cleavage site: 24
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
bacterial outside --- Certainty= 0.300(Affirmative) < succ>
bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3021

A DNA sequence <SEQ ID 9213> was identified in GAS which encodes amino acid sequence <SEQ ID 9214>. Analysis of the amino acid sequence reveals the following:

```

Possible cleavage site: 23
>>> May be a lipoprotein

----- Final Results -----
bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
bacterial outside --- Certainty= 0.000(Not Clear) < succ>
bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3022

A DNA sequence <SEQ ID 9215> was identified in GAS which encodes amino acid sequence <SEQ ID 9216>. Analysis of the amino acid sequence reveals the following:

```

Possible site: 19

>>> Seems to have an uncleavable N-term signal seq
INTEGRAL    Likelihood = -2.76    Transmembrane    3 - 19 ( 2 - 20)

----- Final Results -----
bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

RGD motif: 396-398

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3023

A DNA sequence <SEQ ID 9217> was identified in GAS which encodes amino acid sequence <SEQ ID 9218>. Analysis of the amino acid sequence reveals the following:

-2920-

Possible site: 18

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.80 Transmembrane 251 - 267 (251 - 267)

INTEGRAL Likelihood = -0.75 Transmembrane 179 - 195 (179 - 195)

----- Final Results -----

bacterial membrane --- Certainty=0.1319(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3024

A DNA sequence <SEQ ID 9219> was identified in GAS which encodes amino acid sequence <SEQ ID 9220>. Analysis of the amino acid sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.22 Transmembrane 52 - 68 (51 - 68)

----- Final Results -----

bacterial membrane --- Certainty=0.1489(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3025

A DNA sequence <SEQ ID 9221> was identified in GAS which encodes amino acid sequence <SEQ ID 9222>. Analysis of the amino acid sequence reveals the following:

Possible site: 52

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -12.58 Transmembrane 39 - 55 (32 - 86)

INTEGRAL Likelihood = -9.55 Transmembrane 60 - 76 (56 - 86)

----- Final Results -----

bacterial membrane --- Certainty=0.6031(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3026

A DNA sequence <SEQ ID 9223> was identified in GAS which encodes amino acid sequence <SEQ ID 9224>. Analysis of the amino acid sequence reveals the following:

Possible site: 18

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2921-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

5 Example 3027

A DNA sequence <SEQ ID 9225> was identified in GAS which encodes amino acid sequence <SEQ ID 9226>. Analysis of the amino acid sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3028

A DNA sequence <SEQ ID 9227> was identified in GAS which encodes amino acid sequence <SEQ ID 9228>. Analysis of the amino acid sequence reveals the following:

Possible site: 33
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -8.44 Transmembrane 18 - 34 (13 - 40)
 INTEGRAL Likelihood = -7.86 Transmembrane 59 - 75 (54 - 79)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4376 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3029

A DNA sequence <SEQ ID 9229> was identified in GAS which encodes amino acid sequence <SEQ ID 9230>. Analysis of the amino acid sequence reveals the following:

Possible site: 27
 >>> Seems to have a cleavable N-term signal seq.
 ----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3030

A DNA sequence <SEQ ID 9231> was identified in GAS which encodes amino acid sequence <SEQ ID 9232>. Analysis of the amino acid sequence reveals the following:

-2922-

Possible site: 24
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for
 10 vaccines and/or diagnostics.

Example 3031

A DNA sequence <SEQ ID 9233> was identified in GAS which encodes amino acid sequence <SEQ ID 9234>. Analysis of the amino acid sequence reveals the following:

Possible site: 49
 15 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -9.87 Transmembrane 58 - 74 (53 - 81)

----- Final Results -----

20 bacterial membrane --- Certainty=0.4949(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for
 vaccines and/or diagnostics.

25 Example 3032

A DNA sequence <SEQ ID 9235> was identified in GAS which encodes amino acid sequence <SEQ ID 9236>. Analysis of the amino acid sequence reveals the following:

Possible site: 16
 30 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.06 Transmembrane 92 - 108 (92 - 108)

----- Final Results -----

35 bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for
 vaccines and/or diagnostics.

Example 3033

40 A DNA sequence <SEQ ID 9237> was identified in GAS which encodes amino acid sequence <SEQ ID 9238>. Analysis of the amino acid sequence reveals the following:

Possible site: 40
 45 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.38 Transmembrane 18 - 34 (18 - 34)

----- Final Results -----

50 bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for
 vaccines and/or diagnostics.

Example 3034

A DNA sequence <SEQ ID 9239> was identified in GAS which encodes amino acid sequence <SEQ ID 9240>. Analysis of the amino acid sequence reveals the following:

```

5      Possible site: 19
      >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
10     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3035

15 A DNA sequence <SEQ ID 9241> was identified in GAS which encodes amino acid sequence <SEQ ID 9242>. Analysis of the amino acid sequence reveals the following:

```

      Possible site: 57
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.01    Transmembrane  155 - 171 ( 154 - 171)
20
      ----- Final Results -----
      bacterial membrane --- Certainty=0.1404(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
25     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3036

30 A DNA sequence <SEQ ID 9243> was identified in GAS which encodes amino acid sequence <SEQ ID 9244>. Analysis of the amino acid sequence reveals the following:

```

      Possible site: 28
      >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -4.25    Transmembrane  113 - 129 ( 111 - 131)
35
      ----- Final Results -----
      bacterial membrane --- Certainty=0.2699(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

40 Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3037

A DNA sequence <SEQ ID 9245> was identified in GAS which encodes amino acid sequence <SEQ ID 9246>. Analysis of the amino acid sequence reveals the following:

```

45     Possible site: 56
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.69    Transmembrane  110 - 126 ( 110 - 126)

      ----- Final Results -----
50     bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>

```

-2924-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

5 Example 3038

A DNA sequence <SEQ ID 9247> was identified in GAS which encodes amino acid sequence <SEQ ID 9248>. Analysis of the amino acid sequence reveals the following:

```

Possible site: 58
>>> Seems to have no N-terminal signal sequence
10    INTEGRAL    Likelihood = -1.28    Transmembrane    130 - 146 ( 128 - 146)

----- Final Results -----
        bacterial membrane --- Certainty=0.1510 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3039

20 A DNA sequence <SEQ ID 9249> was identified in GAS which encodes amino acid sequence <SEQ ID 9250>. Analysis of the amino acid sequence reveals the following:

```

Possible site: 39
>>> Seems to have no N-terminal signal sequence
25    INTEGRAL    Likelihood = -4.57    Transmembrane    74 - 90 ( 72 - 92)
    INTEGRAL    Likelihood = -3.13    Transmembrane    169 - 185 ( 166 - 185)
    INTEGRAL    Likelihood = -3.13    Transmembrane    28 - 44 ( 27 - 44)

----- Final Results -----
        bacterial membrane --- Certainty=0.2826 (Affirmative) < succ>
30    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

35 Example 3040

A DNA sequence <SEQ ID 9251> was identified in GAS which encodes amino acid sequence <SEQ ID 9252>. Analysis of the amino acid sequence reveals the following:

```

Possible cleavage site: 56
>>> Seems to have a cleavable N-term signal seq.
40    INTEGRAL    Likelihood = -12.21    Transmembrane    93 - 109 ( 87 - 114)
    INTEGRAL    Likelihood = -8.65    Transmembrane    227 - 243 ( 226 - 243)

----- Final Results -----
        bacterial membrane --- Certainty=0.588 (Affirmative) < succ>
45    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

-2925-

Example 3041

A DNA sequence <SEQ ID 9253> was identified in GAS which encodes amino acid sequence <SEQ ID 9254>. Analysis of the amino acid sequence reveals the following:

```

Possible site: 45
5  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -6.53    Transmembrane    73 - 89 ( 70 - 94)
    INTEGRAL    Likelihood = -4.41    Transmembrane    32 - 48 ( 30 - 51)
    INTEGRAL    Likelihood = -2.55    Transmembrane    10 - 26 ( 10 - 26)
10  INTEGRAL    Likelihood = -2.39    Transmembrane    106 - 122 ( 104 - 123)
    INTEGRAL    Likelihood = -1.75    Transmembrane    153 - 169 ( 152 - 169)

----- Final Results -----
    bacterial membrane --- Certainty=0.3612(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3042

20 A DNA sequence <SEQ ID 9255> was identified in GAS which encodes amino acid sequence <SEQ ID 9256>. Analysis of the amino acid sequence reveals the following:

```

Possible site: 44
    >>> Seems to have no N-terminal signal sequence
25  INTEGRAL    Likelihood = -11.68    Transmembrane    25 - 41 ( 15 - 46)
    INTEGRAL    Likelihood = -9.24    Transmembrane    255 - 271 ( 248 - 276)
    INTEGRAL    Likelihood = -7.59    Transmembrane    82 - 98 ( 79 - 100)
    INTEGRAL    Likelihood = -4.30    Transmembrane    115 - 131 ( 113 - 135)
    INTEGRAL    Likelihood = -0.11    Transmembrane    148 - 164 ( 148 - 164)

30  ----- Final Results -----
    bacterial membrane --- Certainty=0.5670(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

35 Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3043

A DNA sequence <SEQ ID 9257> was identified in GAS which encodes amino acid sequence <SEQ ID 9258>. Analysis of the amino acid sequence reveals the following:

```

40  Possible site: 51
    >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -11.04    Transmembrane    137 - 153 ( 126 - 160)
    INTEGRAL    Likelihood = -10.56    Transmembrane    36 - 52 ( 29 - 58)
    INTEGRAL    Likelihood = -10.08    Transmembrane    407 - 423 ( 399 - 426)
45  INTEGRAL    Likelihood = -4.94    Transmembrane    230 - 246 ( 228 - 250)
    INTEGRAL    Likelihood = -4.83    Transmembrane    79 - 95 ( 77 - 98)
    INTEGRAL    Likelihood = -4.35    Transmembrane    202 - 218 ( 201 - 220)
    INTEGRAL    Likelihood = -1.12    Transmembrane    293 - 309 ( 293 - 309)

50  ----- Final Results -----
    bacterial membrane --- Certainty=0.5416(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-2926-

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3044

A DNA sequence <SEQ ID 9259> was identified in GAS which encodes amino acid sequence <SEQ ID 9260>. Analysis of the amino acid sequence reveals the following:

```

Possible site: 31
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -2.76    Transmembrane  137 - 153 ( 137 - 154)

----- Final Results -----
bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3045

A DNA sequence <SEQ ID 9261> was identified in GAS which encodes amino acid sequence <SEQ ID 9262>. Analysis of the amino acid sequence reveals the following:

```

Possible site: 36
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -7.91    Transmembrane  238 - 254 ( 236 - 264)
INTEGRAL    Likelihood = -6.16    Transmembrane   69 - 85 ( 65 - 89)
INTEGRAL    Likelihood = -6.00    Transmembrane  136 - 152 ( 134 - 155)
INTEGRAL    Likelihood = -4.73    Transmembrane   29 - 45 ( 21 - 48)
INTEGRAL    Likelihood = -2.97    Transmembrane  194 - 210 ( 193 - 220)

----- Final Results -----
bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3046

A DNA sequence <SEQ ID 9263> was identified in GAS which encodes amino acid sequence <SEQ ID 9264>. Analysis of the amino acid sequence reveals the following:

```

Possible site: 39
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL    Likelihood = -9.87    Transmembrane  574 - 590 ( 568 - 601)
INTEGRAL    Likelihood = -9.18    Transmembrane  243 - 259 ( 238 - 262)
INTEGRAL    Likelihood = -7.11    Transmembrane   66 - 82 ( 65 - 87)
INTEGRAL    Likelihood = -1.28    Transmembrane  270 - 286 ( 270 - 287)

----- Final Results -----
bacterial membrane --- Certainty=0.4949(Affirmative) < succ>
bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

-2927-

Example 3047

A DNA sequence <SEQ ID 9265> was identified in GAS which encodes amino acid sequence <SEQ ID 9266>. Analysis of the amino acid sequence reveals the following:

```

5      Possible site: 33
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -7.91    Transmembrane    98 - 114 ( 92 - 124)
      INTEGRAL    Likelihood = -6.21    Transmembrane    19 - 35 ( 14 - 37)
      INTEGRAL    Likelihood = -5.36    Transmembrane    170 - 186 ( 169 - 189)
10     INTEGRAL    Likelihood = -5.15    Transmembrane    147 - 163 ( 136 - 167)
      INTEGRAL    Likelihood = -1.12    Transmembrane    77 - 93 ( 77 - 93)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3048

20 A DNA sequence <SEQ ID 9267> was identified in GAS which encodes amino acid sequence <SEQ ID 9268>. Analysis of the amino acid sequence reveals the following:

```

      Possible site: 47
      >>> Seems to have no N-terminal signal sequence
25     INTEGRAL    Likelihood = -11.94   Transmembrane    27 - 43 ( 19 - 51)
      INTEGRAL    Likelihood = -4.83     Transmembrane    152 - 168 ( 151 - 171)
      INTEGRAL    Likelihood = -4.09     Transmembrane    277 - 293 ( 276 - 294)
      INTEGRAL    Likelihood = -3.82     Transmembrane    195 - 211 ( 193 - 217)
      INTEGRAL    Likelihood = -2.50     Transmembrane    120 - 136 ( 120 - 137)
30     INTEGRAL    Likelihood = -0.64     Transmembrane    81 - 97 ( 81 - 98)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.5776(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
35     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3049

40 A DNA sequence <SEQ ID 9269> was identified in GAS which encodes amino acid sequence <SEQ ID 9270>. Analysis of the amino acid sequence reveals the following:

```

      Possible site: 36
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -8.49     Transmembrane    27 - 43 ( 14 - 50)
      INTEGRAL    Likelihood = -8.17     Transmembrane    58 - 74 ( 52 - 79)
45     INTEGRAL    Likelihood = -7.38     Transmembrane    165 - 181 ( 161 - 193)
      INTEGRAL    Likelihood = -3.66     Transmembrane    247 - 263 ( 246 - 270)
      INTEGRAL    Likelihood = -1.54     Transmembrane    134 - 150 ( 134 - 150)

      ----- Final Results -----
50     bacterial membrane --- Certainty=0.440(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-2928-

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3050

A DNA sequence <SEQ ID 9271> was identified in GAS which encodes amino acid sequence <SEQ ID 9272>. Analysis of the amino acid sequence reveals the following:

```

Possible site: 55
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -14.75    Transmembrane    389 - 405 ( 377 - 413)
    INTEGRAL    Likelihood = -8.44     Transmembrane    31 - 47 ( 29 - 54)
10    INTEGRAL    Likelihood = -7.17     Transmembrane    181 - 197 ( 179 - 205)
    INTEGRAL    Likelihood = -7.01     Transmembrane    339 - 355 ( 326 - 360)
    INTEGRAL    Likelihood = -6.58     Transmembrane    105 - 121 ( 102 - 124)
    INTEGRAL    Likelihood = -5.36     Transmembrane    225 - 241 ( 222 - 244)
    INTEGRAL    Likelihood = -0.43     Transmembrane    139 - 155 ( 139 - 155)
15    INTEGRAL    Likelihood = -0.16     Transmembrane    283 - 299 ( 282 - 300)

----- Final Results -----
        bacterial membrane --- Certainty=0.6901(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
20        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3051

25 A DNA sequence <SEQ ID 9273> was identified in GAS which encodes amino acid sequence <SEQ ID 9274>. Analysis of the amino acid sequence reveals the following:

```

Possible cleavage site: 25
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -5.31     Transmembrane    155 - 171 ( 154 - 174)
30    INTEGRAL    Likelihood = -3.50     Transmembrane    111 - 127 ( 110 - 128)
    INTEGRAL    Likelihood = -2.07     Transmembrane    80 - 96 ( 78 - 96)
    INTEGRAL    Likelihood = -0.90     Transmembrane    57 - 73 ( 57 - 74)

----- Final Results -----
35        bacterial membrane --- Certainty=0.312(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

40 Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3052

A DNA sequence <SEQ ID 9275> was identified in GAS which encodes amino acid sequence <SEQ ID 9276>. Analysis of the amino acid sequence reveals the following:

```

Possible site: 27
45 >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -3.93     Transmembrane    463 - 479 ( 461 - 480)

----- Final Results -----
        bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
50        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it is predicted that this GAS protein , and its epitopes, could be useful antigens for vaccines and/or diagnostics.

Example 3053

5 A DNA sequence <SEQ ID 8741> was identified in GBS which encodes amino acid sequence <SEQ ID 8742>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3054

A DNA sequence <SEQ ID 8685> was identified in GBS which encodes amino acid sequence <SEQ ID 8686>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3055

10 A DNA sequence <SEQ ID 10303> was identified in GBS which encodes amino acid sequence <SEQ ID 10304>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3056

A DNA sequence <SEQ ID 10305> was identified in GBS which encodes amino acid sequence <SEQ ID 10306>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

15 **Example 3057**

A DNA sequence <SEQ ID 10307> was identified in GBS which encodes amino acid sequence <SEQ ID 10308>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3058

20 A DNA sequence <SEQ ID 10309> was identified in GBS which encodes amino acid sequence <SEQ ID 10310>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3059

A DNA sequence <SEQ ID 10311> was identified in GBS which encodes amino acid sequence <SEQ ID 10312>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3060

25 A DNA sequence <SEQ ID 10313> was identified in GBS which encodes amino acid sequence <SEQ ID 10314>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3061

A DNA sequence <SEQ ID 10315> was identified in GBS which encodes amino acid sequence <SEQ ID 10316>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

30 **Example 3062**

A DNA sequence <SEQ ID 10317> was identified in GBS which encodes amino acid sequence <SEQ ID 10318>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3063

35 A repeated DNA sequence <SEQ ID 10319> was identified in GBS which encodes amino acid sequence <SEQ ID 10320>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3064

A DNA sequence <SEQ ID 10321> was identified in GBS which encodes amino acid sequence <SEQ ID 10322>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3065

- 5 A DNA sequence <SEQ ID 10323> was identified in GBS which encodes amino acid sequence <SEQ ID 10324>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3066

A DNA sequence <SEQ ID 10325> was identified in GBS which encodes amino acid sequence <SEQ ID 10326>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3067**

A DNA sequence <SEQ ID 10327> was identified in GBS which encodes amino acid sequence <SEQ ID 10328>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3068

- 15 A DNA sequence <SEQ ID 10329> was identified in GBS which encodes amino acid sequence <SEQ ID 10330>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3069

A DNA sequence <SEQ ID 10331> was identified in GBS which encodes amino acid sequence <SEQ ID 10332>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3070

- 20 A DNA sequence <SEQ ID 10333> was identified in GBS which encodes amino acid sequence <SEQ ID 10334>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3071

A DNA sequence <SEQ ID 10335> was identified in GBS which encodes amino acid sequence <SEQ ID 10336>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3072**

A DNA sequence <SEQ ID 10339> was identified in GBS which encodes amino acid sequence <SEQ ID 10340>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3073

- 30 A DNA sequence <SEQ ID 10341> was identified in GBS which encodes amino acid sequence <SEQ ID 10342>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3074

A DNA sequence <SEQ ID 10343> was identified in GBS which encodes amino acid sequence <SEQ ID 10344>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3075

- 35 A DNA sequence <SEQ ID 10345> was identified in GBS which encodes amino acid sequence <SEQ ID 10346>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3076

A DNA sequence <SEQ ID 10347> was identified in GBS which encodes amino acid sequence <SEQ ID 10348>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3077

- 5 A DNA sequence <SEQ ID 10349> was identified in GBS which encodes amino acid sequence <SEQ ID 10350>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3078

A DNA sequence <SEQ ID 10351> was identified in GBS which encodes amino acid sequence <SEQ ID 10352>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3079**

A DNA sequence <SEQ ID 10353> was identified in GBS which encodes amino acid sequence <SEQ ID 10354>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3080

- 15 A DNA sequence <SEQ ID 10355> was identified in GBS which encodes amino acid sequence <SEQ ID 10356>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3081

A DNA sequence <SEQ ID 10357> was identified in GBS which encodes amino acid sequence <SEQ ID 10358>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3082

- 20 A DNA sequence <SEQ ID 10359> was identified in GBS which encodes amino acid sequence <SEQ ID 10360>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3083

A DNA sequence <SEQ ID 10361> was identified in GBS which encodes amino acid sequence <SEQ ID 10362>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3084**

A DNA sequence <SEQ ID 10363> was identified in GBS which encodes amino acid sequence <SEQ ID 10364>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3085

- 30 A DNA sequence <SEQ ID 10365> was identified in GBS which encodes amino acid sequence <SEQ ID 10366>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3086

A DNA sequence <SEQ ID 10367> was identified in GBS which encodes amino acid sequence <SEQ ID 10368>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3087

- 35 A DNA sequence <SEQ ID 10369> was identified in GBS which encodes amino acid sequence <SEQ ID 10370>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3088

A DNA sequence <SEQ ID 10371> was identified in GBS which encodes amino acid sequence <SEQ ID 10372>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3089

- 5 A DNA sequence <SEQ ID 10373> was identified in GBS which encodes amino acid sequence <SEQ ID 10374>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3090

A DNA sequence <SEQ ID 10375> was identified in GBS which encodes amino acid sequence <SEQ ID 10376>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3091**

A DNA sequence <SEQ ID 10377> was identified in GBS which encodes amino acid sequence <SEQ ID 10378>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3092

- 15 A DNA sequence <SEQ ID 10379> was identified in GBS which encodes amino acid sequence <SEQ ID 10380>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3093

A DNA sequence <SEQ ID 10381> was identified in GBS which encodes amino acid sequence <SEQ ID 10382>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3094

- 20 A DNA sequence <SEQ ID 10383> was identified in GBS which encodes amino acid sequence <SEQ ID 10384>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3095

A DNA sequence <SEQ ID 10385> was identified in GBS which encodes amino acid sequence <SEQ ID 10386>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3096**

A DNA sequence <SEQ ID 10387> was identified in GBS which encodes amino acid sequence <SEQ ID 10388>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3097

- 30 A DNA sequence <SEQ ID 10389> was identified in GBS which encodes amino acid sequence <SEQ ID 10390>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3098

A DNA sequence <SEQ ID 10391> was identified in GBS which encodes amino acid sequence <SEQ ID 10392>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3099

- 35 A DNA sequence <SEQ ID 10393> was identified in GBS which encodes amino acid sequence <SEQ ID 10394>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3100

A DNA sequence <SEQ ID 10395> was identified in GBS which encodes amino acid sequence <SEQ ID 10396>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3101

- 5 A DNA sequence <SEQ ID 10397> was identified in GBS which encodes amino acid sequence <SEQ ID 10398>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3102

A DNA sequence <SEQ ID 10399> was identified in GBS which encodes amino acid sequence <SEQ ID 10400>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3103**

A DNA sequence <SEQ ID 10401> was identified in GBS which encodes amino acid sequence <SEQ ID 10402>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3104

- 15 A DNA sequence <SEQ ID 10403> was identified in GBS which encodes amino acid sequence <SEQ ID 10404>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3105

A DNA sequence <SEQ ID 10405> was identified in GBS which encodes amino acid sequence <SEQ ID 10406>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3106

- 20 A DNA sequence <SEQ ID 10407> was identified in GBS which encodes amino acid sequence <SEQ ID 10408>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3107

A DNA sequence <SEQ ID 10409> was identified in GBS which encodes amino acid sequence <SEQ ID 10410>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3108**

A DNA sequence <SEQ ID 10411> was identified in GBS which encodes amino acid sequence <SEQ ID 10412>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3109

- 30 A DNA sequence <SEQ ID 10413> was identified in GBS which encodes amino acid sequence <SEQ ID 10414>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3110

A DNA sequence <SEQ ID 10415> was identified in GBS which encodes amino acid sequence <SEQ ID 10416>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3111

- 35 A DNA sequence <SEQ ID 10417> was identified in GBS which encodes amino acid sequence <SEQ ID 10418>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3112

A DNA sequence <SEQ ID 10419> was identified in GBS which encodes amino acid sequence <SEQ ID 10420>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3113

- 5 A DNA sequence <SEQ ID 10421> was identified in GBS which encodes amino acid sequence <SEQ ID 10422>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3114

A DNA sequence <SEQ ID 10423> was identified in GBS which encodes amino acid sequence <SEQ ID 10424>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3115**

A DNA sequence <SEQ ID 10425> was identified in GBS which encodes amino acid sequence <SEQ ID 10426>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3116

- 15 A DNA sequence <SEQ ID 10427> was identified in GBS which encodes amino acid sequence <SEQ ID 10428>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3117

A DNA sequence <SEQ ID 10429> was identified in GBS which encodes amino acid sequence <SEQ ID 10430>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3118

- 20 A DNA sequence <SEQ ID 10431> was identified in GBS which encodes amino acid sequence <SEQ ID 10432>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3119

A DNA sequence <SEQ ID 10433> was identified in GBS which encodes amino acid sequence <SEQ ID 10434>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3120**

A DNA sequence <SEQ ID 10435> was identified in GBS which encodes amino acid sequence <SEQ ID 10436>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3121

- 30 A DNA sequence <SEQ ID 10437> was identified in GBS which encodes amino acid sequence <SEQ ID 10438>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3122

A DNA sequence <SEQ ID 10441> was identified in GBS which encodes amino acid sequence <SEQ ID 10442>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3123

- 35 A DNA sequence <SEQ ID 10443> was identified in GBS which encodes amino acid sequence <SEQ ID 10444>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3124

A DNA sequence <SEQ ID 10445> was identified in GBS which encodes amino acid sequence <SEQ ID 10446>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3125

- 5 A DNA sequence <SEQ ID 10447> was identified in GBS which encodes amino acid sequence <SEQ ID 10448>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3126

A DNA sequence <SEQ ID 10449> was identified in GBS which encodes amino acid sequence <SEQ ID 10450>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3127**

A DNA sequence <SEQ ID 10451> was identified in GBS which encodes amino acid sequence <SEQ ID 10452>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3128

- 15 A DNA sequence <SEQ ID 10453> was identified in GBS which encodes amino acid sequence <SEQ ID 10454>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3129

A DNA sequence <SEQ ID 10455> was identified in GBS which encodes amino acid sequence <SEQ ID 10456>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3130

- 20 A DNA sequence <SEQ ID 10457> was identified in GBS which encodes amino acid sequence <SEQ ID 10458>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics. A related GBS nucleic acid sequence <SEQ ID 10907> which encodes amino acid sequence <SEQ ID 10908> was also identified.

Example 3131

- 25 A DNA sequence <SEQ ID 10459> was identified in GBS which encodes amino acid sequence <SEQ ID 10460>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3132

A DNA sequence <SEQ ID 10461> was identified in GBS which encodes amino acid sequence <SEQ ID 10462>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

30 **Example 3133**

A DNA sequence <SEQ ID 10463> was identified in GBS which encodes amino acid sequence <SEQ ID 10464>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3134

- 35 A DNA sequence <SEQ ID 10465> was identified in GBS which encodes amino acid sequence <SEQ ID 10466>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3135

A DNA sequence <SEQ ID 10467> was identified in GBS which encodes amino acid sequence <SEQ ID 10468>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3136

- 5 A DNA sequence <SEQ ID 10469> was identified in GBS which encodes amino acid sequence <SEQ ID 10470>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3137

A DNA sequence <SEQ ID 10471> was identified in GBS which encodes amino acid sequence <SEQ ID 10472>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3138**

A DNA sequence <SEQ ID 10473> was identified in GBS which encodes amino acid sequence <SEQ ID 10474>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3139

- 15 A DNA sequence <SEQ ID 10475> was identified in GBS which encodes amino acid sequence <SEQ ID 10476>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3140

A DNA sequence <SEQ ID 10477> was identified in GBS which encodes amino acid sequence <SEQ ID 10478>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3141

- 20 A DNA sequence <SEQ ID 10479> was identified in GBS which encodes amino acid sequence <SEQ ID 10480>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3142

A DNA sequence <SEQ ID 10481> was identified in GBS which encodes amino acid sequence <SEQ ID 10482>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3143**

A DNA sequence <SEQ ID 10483> was identified in GBS which encodes amino acid sequence <SEQ ID 10484>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3144

- 30 A DNA sequence <SEQ ID 10485> was identified in GBS which encodes amino acid sequence <SEQ ID 10486>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3145

A DNA sequence <SEQ ID 10487> was identified in GBS which encodes amino acid sequence <SEQ ID 10488>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3146

- 35 A DNA sequence <SEQ ID 10489> was identified in GBS which encodes amino acid sequence <SEQ ID 10490>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3147

A DNA sequence <SEQ ID 10491> was identified in GBS which encodes amino acid sequence <SEQ ID 10492>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3148

- 5 A DNA sequence <SEQ ID 10493> was identified in GBS which encodes amino acid sequence <SEQ ID 10494>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3149

A DNA sequence <SEQ ID 10495> was identified in GBS which encodes amino acid sequence <SEQ ID 10496>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3150**

A DNA sequence <SEQ ID 10497> was identified in GBS which encodes amino acid sequence <SEQ ID 10498>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3151

- 15 A DNA sequence <SEQ ID 10499> was identified in GBS which encodes amino acid sequence <SEQ ID 10500>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3152

A DNA sequence <SEQ ID 10501> was identified in GBS which encodes amino acid sequence <SEQ ID 10502>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3153

- 20 A DNA sequence <SEQ ID 10503> was identified in GBS which encodes amino acid sequence <SEQ ID 10504>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3154

A DNA sequence <SEQ ID 10505> was identified in GBS which encodes amino acid sequence <SEQ ID 10506>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3155**

A DNA sequence <SEQ ID 10509> was identified in GBS which encodes amino acid sequence <SEQ ID 10510>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3156

- 30 A DNA sequence <SEQ ID 10511> was identified in GBS which encodes amino acid sequence <SEQ ID 10512>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3157

A DNA sequence <SEQ ID 10513> was identified in GBS which encodes amino acid sequence <SEQ ID 10514>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3158

- 35 A DNA sequence <SEQ ID 10515> was identified in GBS which encodes amino acid sequence <SEQ ID 10516>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3159

A DNA sequence <SEQ ID 10517> was identified in GBS which encodes amino acid sequence <SEQ ID 10518>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3160

- 5 A DNA sequence <SEQ ID 10519> was identified in GBS which encodes amino acid sequence <SEQ ID 10520>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3161

A DNA sequence <SEQ ID 10521> was identified in GBS which encodes amino acid sequence <SEQ ID 10522>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3162**

A DNA sequence <SEQ ID 10523> was identified in GBS which encodes amino acid sequence <SEQ ID 10524>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3163

- 15 A DNA sequence <SEQ ID 10525> was identified in GBS which encodes amino acid sequence <SEQ ID 10526>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3164

A DNA sequence <SEQ ID 10527> was identified in GBS which encodes amino acid sequence <SEQ ID 10528>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3165

- 20 A DNA sequence <SEQ ID 10529> was identified in GBS which encodes amino acid sequence <SEQ ID 10530>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3166

A DNA sequence <SEQ ID 10531> was identified in GBS which encodes amino acid sequence <SEQ ID 10532>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3167**

A DNA sequence <SEQ ID 10533> was identified in GBS which encodes amino acid sequence <SEQ ID 10534>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3168

- 30 A DNA sequence <SEQ ID 10535> was identified in GBS which encodes amino acid sequence <SEQ ID 10536>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3169

A DNA sequence <SEQ ID 10537> was identified in GBS which encodes amino acid sequence <SEQ ID 10538>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3170

- 35 A DNA sequence <SEQ ID 10539> was identified in GBS which encodes amino acid sequence <SEQ ID 10540>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3171

A DNA sequence <SEQ ID 10541> was identified in GBS which encodes amino acid sequence <SEQ ID 10542>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3172

- 5 A DNA sequence <SEQ ID 10543> was identified in GBS which encodes amino acid sequence <SEQ ID 10544>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3173

A DNA sequence <SEQ ID 10545> was identified in GBS which encodes amino acid sequence <SEQ ID 10546>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

- 10 SEQ ID 10546 (GBS665) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 137 (lane 8-10; MW 41kDa) and in Figure 187 (lane 5; MW 41kDa). It was also was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 137 (lane 11 & 12; MW 16.1kDa), in Figure 141 (lane 4; MW 16kDa) and in Figure 179 (lane 6; MW 16kDa). Purified GBS665-GST is shown in Figure 243, lane 4.
- 15 GBS665-His was purified as shown in Figure 230, lane 7-8.

Example 3174

- A DNA sequence <SEQ ID 10547> was identified in GBS which encodes amino acid sequence <SEQ ID 10548>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics. A related GBS nucleic acid sequence <SEQ ID 10909> which encodes amino acid sequence <SEQ ID 10910> was
- 20 also identified.

Example 3175

A DNA sequence <SEQ ID 10549> was identified in GBS which encodes amino acid sequence <SEQ ID 10550>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3176

- 25 A DNA sequence <SEQ ID 10551> was identified in GBS which encodes amino acid sequence <SEQ ID 10552>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3177

A DNA sequence <SEQ ID 10553> was identified in GBS which encodes amino acid sequence <SEQ ID 10554>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

- 30 **Example 3178**

A DNA sequence <SEQ ID 10555> was identified in GBS which encodes amino acid sequence <SEQ ID 10556>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3179

- A DNA sequence <SEQ ID 10557> was identified in GBS which encodes amino acid sequence <SEQ ID
- 35 10558>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3180

A DNA sequence <SEQ ID 10559> was identified in GBS which encodes amino acid sequence <SEQ ID 10560>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3181

- 5 A DNA sequence <SEQ ID 10561> was identified in GBS which encodes amino acid sequence <SEQ ID 10562>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3182

A DNA sequence <SEQ ID 10563> was identified in GBS which encodes amino acid sequence <SEQ ID 10564>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3183**

A DNA sequence <SEQ ID 10565> was identified in GBS which encodes amino acid sequence <SEQ ID 10566>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3184

- 15 A DNA sequence <SEQ ID 10567> was identified in GBS which encodes amino acid sequence <SEQ ID 10568>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3185

A DNA sequence <SEQ ID 10569> was identified in GBS which encodes amino acid sequence <SEQ ID 10570>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3186

- 20 A DNA sequence <SEQ ID 10571> was identified in GBS which encodes amino acid sequence <SEQ ID 10572>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3187

A DNA sequence <SEQ ID 10573> was identified in GBS which encodes amino acid sequence <SEQ ID 10574>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3188**

A DNA sequence <SEQ ID 10575> was identified in GBS which encodes amino acid sequence <SEQ ID 10576>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3189

- 30 A DNA sequence <SEQ ID 10577> was identified in GBS which encodes amino acid sequence <SEQ ID 10578>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3190

A DNA sequence <SEQ ID 10579> was identified in GBS which encodes amino acid sequence <SEQ ID 10580>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3191

- 35 A DNA sequence <SEQ ID 10581> was identified in GBS which encodes amino acid sequence <SEQ ID 10582>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3192

A DNA sequence <SEQ ID 10583> was identified in GBS which encodes amino acid sequence <SEQ ID 10584>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3193

- 5 A DNA sequence <SEQ ID 10585> was identified in GBS which encodes amino acid sequence <SEQ ID 10586>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3194

A DNA sequence <SEQ ID 10587> was identified in GBS which encodes amino acid sequence <SEQ ID 10588>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3195**

A DNA sequence <SEQ ID 10591> was identified in GBS which encodes amino acid sequence <SEQ ID 10592>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3196

- 15 A DNA sequence <SEQ ID 10593> was identified in GBS which encodes amino acid sequence <SEQ ID 10594>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3197

A DNA sequence <SEQ ID 10595> was identified in GBS which encodes amino acid sequence <SEQ ID 10596>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3198

- 20 A DNA sequence <SEQ ID 10597> was identified in GBS which encodes amino acid sequence <SEQ ID 10598>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics. A related GBS nucleic acid sequence <SEQ ID 10797> which encodes amino acid sequence <SEQ ID 10798> was also identified.

Example 3199

- 25 A DNA sequence <SEQ ID 10599> was identified in GBS which encodes amino acid sequence <SEQ ID 10600>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3200

A DNA sequence <SEQ ID 10601> was identified in GBS which encodes amino acid sequence <SEQ ID 10602>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

30 **Example 3201**

A DNA sequence <SEQ ID 10603> was identified in GBS which encodes amino acid sequence <SEQ ID 10604>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3202

- 35 A DNA sequence <SEQ ID 10605> was identified in GBS which encodes amino acid sequence <SEQ ID 10606>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3203

A DNA sequence <SEQ ID 10607> was identified in GBS which encodes amino acid sequence <SEQ ID 10608>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3204

- 5 A DNA sequence <SEQ ID 10609> was identified in GBS which encodes amino acid sequence <SEQ ID 10610>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3205

A DNA sequence <SEQ ID 10611> was identified in GBS which encodes amino acid sequence <SEQ ID 10612>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3206**

A DNA sequence <SEQ ID 10613> was identified in GBS which encodes amino acid sequence <SEQ ID 10614>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3207

- 15 A DNA sequence <SEQ ID 10615> was identified in GBS which encodes amino acid sequence <SEQ ID 10616>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3208

A DNA sequence <SEQ ID 10617> was identified in GBS which encodes amino acid sequence <SEQ ID 10618>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3209

- 20 A DNA sequence <SEQ ID 10619> was identified in GBS which encodes amino acid sequence <SEQ ID 10620>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3210

A DNA sequence <SEQ ID 10621> was identified in GBS which encodes amino acid sequence <SEQ ID 10622>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3211**

A DNA sequence <SEQ ID 10623> was identified in GBS which encodes amino acid sequence <SEQ ID 10624>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3212

- 30 A DNA sequence <SEQ ID 10625> was identified in GBS which encodes amino acid sequence <SEQ ID 10626>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3213

A DNA sequence <SEQ ID 10627> was identified in GBS which encodes amino acid sequence <SEQ ID 10628>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3214

- 35 A DNA sequence <SEQ ID 10629> was identified in GBS which encodes amino acid sequence <SEQ ID 10630>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3215

A DNA sequence <SEQ ID 10631> was identified in GBS which encodes amino acid sequence <SEQ ID 10632>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3216

- 5 A DNA sequence <SEQ ID 10633> was identified in GBS which encodes amino acid sequence <SEQ ID 10634>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics. A related GBS nucleic acid sequence <SEQ ID 10939> which encodes amino acid sequence <SEQ ID 10940> was also identified.

10 SEQ ID 10634 (GBS675) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 162 (lane 14 & 15; MW 56kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 163 (lane 2; MW 31kDa) and in Figure 188 (lane 5; MW 31kDa).

Purified GBS675-His is shown in Figure 240, lane 7-8.

Example 3217

- 15 A DNA sequence <SEQ ID 10635> was identified in GBS which encodes amino acid sequence <SEQ ID 10636>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3218

A DNA sequence <SEQ ID 10637> was identified in GBS which encodes amino acid sequence <SEQ ID 10638>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

20 **Example 3219**

A DNA sequence <SEQ ID 10639> was identified in GBS which encodes amino acid sequence <SEQ ID 10640>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3220

- 25 A DNA sequence <SEQ ID 10641> was identified in GBS which encodes amino acid sequence <SEQ ID 10642>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3221

A DNA sequence <SEQ ID 10643> was identified in GBS which encodes amino acid sequence <SEQ ID 10644>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3222

- 30 A DNA sequence <SEQ ID 10645> was identified in GBS which encodes amino acid sequence <SEQ ID 10646>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3223

A DNA sequence <SEQ ID 10647> was identified in GBS which encodes amino acid sequence <SEQ ID 10648>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

35 **Example 3224**

A DNA sequence <SEQ ID 10649> was identified in GBS which encodes amino acid sequence <SEQ ID 10650>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3225

A DNA sequence <SEQ ID 10651> was identified in GBS which encodes amino acid sequence <SEQ ID 10652>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3226

- 5 A DNA sequence <SEQ ID 10653> was identified in GBS which encodes amino acid sequence <SEQ ID 10654>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3227

A DNA sequence <SEQ ID 10655> was identified in GBS which encodes amino acid sequence <SEQ ID 10656>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3228**

A DNA sequence <SEQ ID 10657> was identified in GBS which encodes amino acid sequence <SEQ ID 10658>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3229

- 15 A DNA sequence <SEQ ID 10659> was identified in GBS which encodes amino acid sequence <SEQ ID 10660>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3230

A DNA sequence <SEQ ID 10661> was identified in GBS which encodes amino acid sequence <SEQ ID 10662>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3231

- 20 A DNA sequence <SEQ ID 10663> was identified in GBS which encodes amino acid sequence <SEQ ID 10664>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3232

- 25 A DNA sequence <SEQ ID 10665> was identified in GBS which encodes amino acid sequence <SEQ ID 10666>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics. A related GBS nucleic acid sequence <SEQ ID 10917> which encodes amino acid sequence <SEQ ID 10918> was also identified.

A DNA sequence <SEQ ID 10667> was identified in GBS which encodes amino acid sequence <SEQ ID 10668>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3233

- 30 A DNA sequence <SEQ ID 10669> was identified in GBS which encodes amino acid sequence <SEQ ID 10670>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3234

A DNA sequence <SEQ ID 10671> was identified in GBS which encodes amino acid sequence <SEQ ID 10672>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

35 **Example 3235**

A DNA sequence <SEQ ID 10673> was identified in GBS which encodes amino acid sequence <SEQ ID 10674>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3236

A DNA sequence <SEQ ID 10675> was identified in GBS which encodes amino acid sequence <SEQ ID 10676>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3237

- 5 A DNA sequence <SEQ ID 10677> was identified in GBS which encodes amino acid sequence <SEQ ID 10678>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3238

A DNA sequence <SEQ ID 10679> was identified in GBS which encodes amino acid sequence <SEQ ID 10680>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3239**

A DNA sequence <SEQ ID 10681> was identified in GBS which encodes amino acid sequence <SEQ ID 10682>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3240

- 15 A DNA sequence <SEQ ID 10683> was identified in GBS which encodes amino acid sequence <SEQ ID 10684>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3241

A DNA sequence <SEQ ID 10685> was identified in GBS which encodes amino acid sequence <SEQ ID 10686>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3242

- 20 A DNA sequence <SEQ ID 10687> was identified in GBS which encodes amino acid sequence <SEQ ID 10688>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3243

A DNA sequence <SEQ ID 10689> was identified in GBS which encodes amino acid sequence <SEQ ID 10690>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3244**

A DNA sequence <SEQ ID 10691> was identified in GBS which encodes amino acid sequence <SEQ ID 10692>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

- 30 SEQ ID 10692 (GBS676) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 163 (lane 3-5; MW 66kDa) and in Figure 239 (lane 8; MW 66kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 163 (lane 7 & 8; MW 41kDa) and in Figure 188 (lane 6; MW 41kDa). Purified GBS676-His is shown in Figure 240, lane 4-5. Purified GBS676-GST is shown in Figure 246, lanes 10 & 11.

Example 3245

- 35 A DNA sequence <SEQ ID 10693> was identified in GBS which encodes amino acid sequence <SEQ ID 10694>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3246

A DNA sequence <SEQ ID 10695> was identified in GBS which encodes amino acid sequence <SEQ ID 10696>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3247

- 5 A DNA sequence <SEQ ID 10697> was identified in GBS which encodes amino acid sequence <SEQ ID 10698>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3248

A DNA sequence <SEQ ID 10699> was identified in GBS which encodes amino acid sequence <SEQ ID 10700>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3249**

A DNA sequence <SEQ ID 10703> was identified in GBS which encodes amino acid sequence <SEQ ID 10704>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3250

- 15 A DNA sequence <SEQ ID 10705> was identified in GBS which encodes amino acid sequence <SEQ ID 10706>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3251

A DNA sequence <SEQ ID 10707> was identified in GBS which encodes amino acid sequence <SEQ ID 10708>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3252

- 20 A DNA sequence <SEQ ID 10709> was identified in GBS which encodes amino acid sequence <SEQ ID 10710>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics. A related GBS nucleic acid sequence <SEQ ID 10803> which encodes amino acid sequence <SEQ ID 10804> was also identified.

Example 3253

- 25 A DNA sequence <SEQ ID 10711> was identified in GBS which encodes amino acid sequence <SEQ ID 10712>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics. A related GBS nucleic acid sequence <SEQ ID 10913> which encodes amino acid sequence <SEQ ID 10914> was also identified.

Example 3254

- 30 A DNA sequence <SEQ ID 10713> was identified in GBS which encodes amino acid sequence <SEQ ID 10714>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3255

A DNA sequence <SEQ ID 10715> was identified in GBS which encodes amino acid sequence <SEQ ID 10716>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

35 **Example 3256**

A DNA sequence <SEQ ID 10717> was identified in GBS which encodes amino acid sequence <SEQ ID 10718>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3257

A DNA sequence <SEQ ID 10719> was identified in GBS which encodes amino acid sequence <SEQ ID 10720>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3258

- 5 A DNA sequence <SEQ ID 10721> was identified in GBS which encodes amino acid sequence <SEQ ID 10722>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3259

A DNA sequence <SEQ ID 10723> was identified in GBS which encodes amino acid sequence <SEQ ID 10724>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3260**

A DNA sequence <SEQ ID 10725> was identified in GBS which encodes amino acid sequence <SEQ ID 10726>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3261

- 15 A DNA sequence <SEQ ID 10727> was identified in GBS which encodes amino acid sequence <SEQ ID 10728>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3262

A DNA sequence <SEQ ID 10729> was identified in GBS which encodes amino acid sequence <SEQ ID 10730>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

- 20 SEQ ID 10730 (GBS670) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 140 (lane 2-4; MW 45.3kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 140 (lane 5-7; MW 20.4kDa) and in Figure 179 (lane 10; MW 20kDa).

GBS670-His was purified as shown in Figure 230, lane 9-10.

Example 3263

- 25 A DNA sequence <SEQ ID 10731> was identified in GBS which encodes amino acid sequence <SEQ ID 10732>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3264

A DNA sequence <SEQ ID 10733> was identified in GBS which encodes amino acid sequence <SEQ ID 10734>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

30 **Example 3265**

A DNA sequence <SEQ ID 10735> was identified in GBS which encodes amino acid sequence <SEQ ID 10736>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3266

- 35 A DNA sequence <SEQ ID 10737> was identified in GBS which encodes amino acid sequence <SEQ ID 10738>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3267

A DNA sequence <SEQ ID 10739> was identified in GBS which encodes amino acid sequence <SEQ ID 10740>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3268

- 5 A DNA sequence <SEQ ID 10741> was identified in GBS which encodes amino acid sequence <SEQ ID 10742>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3269

A DNA sequence <SEQ ID 10743> was identified in GBS which encodes amino acid sequence <SEQ ID 10744>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3270**

A DNA sequence <SEQ ID 10745> was identified in GBS which encodes amino acid sequence <SEQ ID 10746>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3271

- 15 A DNA sequence <SEQ ID 10747> was identified in GBS which encodes amino acid sequence <SEQ ID 10748>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3272

A DNA sequence <SEQ ID 10749> was identified in GBS which encodes amino acid sequence <SEQ ID 10750>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3273

- 20 A DNA sequence <SEQ ID 10751> was identified in GBS which encodes amino acid sequence <SEQ ID 10752>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3274

A DNA sequence <SEQ ID 10753> was identified in GBS which encodes amino acid sequence <SEQ ID 10754>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3275**

A DNA sequence <SEQ ID 10755> was identified in GBS which encodes amino acid sequence <SEQ ID 10756>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3276

- 30 A DNA sequence <SEQ ID 10757> was identified in GBS which encodes amino acid sequence <SEQ ID 10758>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3277

A DNA sequence <SEQ ID 10759> was identified in GBS which encodes amino acid sequence <SEQ ID 10760>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3278

- 35 A DNA sequence <SEQ ID 10761> was identified in GBS which encodes amino acid sequence <SEQ ID 10762>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3279

A DNA sequence <SEQ ID 10763> was identified in GBS which encodes amino acid sequence <SEQ ID 10764>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3280

- 5 A DNA sequence <SEQ ID 10765> was identified in GBS which encodes amino acid sequence <SEQ ID 10766>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3281

A DNA sequence <SEQ ID 10767> was identified in GBS which encodes amino acid sequence <SEQ ID 10768>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3282**

A DNA sequence <SEQ ID 10769> was identified in GBS which encodes amino acid sequence <SEQ ID 10770>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3283

- 15 A DNA sequence <SEQ ID 10771> was identified in GBS which encodes amino acid sequence <SEQ ID 10772>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3284

A repeated DNA sequence <SEQ ID 10791> was identified in GBS which encodes amino acid sequence <SEQ ID 10792>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3285

- 20 A DNA sequence <SEQ ID 10805> was identified in GBS which encodes amino acid sequence <SEQ ID 10806>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3286

A DNA sequence <SEQ ID 10807> was identified in GBS which encodes amino acid sequence <SEQ ID 10808>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3287**

A DNA sequence <SEQ ID 10809> was identified in GBS which encodes amino acid sequence <SEQ ID 10810>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3288

- 30 A DNA sequence <SEQ ID 10811> was identified in GBS which encodes amino acid sequence <SEQ ID 10812>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3289

A DNA sequence <SEQ ID 10813> was identified in GBS which encodes amino acid sequence <SEQ ID 10814>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3290

- 35 A DNA sequence <SEQ ID 10815> was identified in GBS which encodes amino acid sequence <SEQ ID 10816>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3291

A DNA sequence <SEQ ID 10817> was identified in GBS which encodes amino acid sequence <SEQ ID 10818>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3292

- 5 A DNA sequence <SEQ ID 10819> was identified in GBS which encodes amino acid sequence <SEQ ID 10820>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3293

A DNA sequence <SEQ ID 10821> was identified in GBS which encodes amino acid sequence <SEQ ID 10822>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3294**

A DNA sequence <SEQ ID 10823> was identified in GBS which encodes amino acid sequence <SEQ ID 10824>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3295

- 15 A DNA sequence <SEQ ID 10825> was identified in GBS which encodes amino acid sequence <SEQ ID 10826>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3296

A DNA sequence <SEQ ID 10827> was identified in GBS which encodes amino acid sequence <SEQ ID 10828>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3297

- 20 A DNA sequence <SEQ ID 10829> was identified in GBS which encodes amino acid sequence <SEQ ID 10830>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3298

A DNA sequence <SEQ ID 10831> was identified in GBS which encodes amino acid sequence <SEQ ID 10832>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3299**

A DNA sequence <SEQ ID 10833> was identified in GBS which encodes amino acid sequence <SEQ ID 10834>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3300

- 30 A DNA sequence <SEQ ID 10835> was identified in GBS which encodes amino acid sequence <SEQ ID 10836>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3301

A DNA sequence <SEQ ID 10837> was identified in GBS which encodes amino acid sequence <SEQ ID 10838>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3302

- 35 A DNA sequence <SEQ ID 10839> was identified in GBS which encodes amino acid sequence <SEQ ID 10840>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3303

A DNA sequence <SEQ ID 10841> was identified in GBS which encodes amino acid sequence <SEQ ID 10842>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3304

- 5 A DNA sequence <SEQ ID 10843> was identified in GBS which encodes amino acid sequence <SEQ ID 10844>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3305

A DNA sequence <SEQ ID 10845> was identified in GBS which encodes amino acid sequence <SEQ ID 10846>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3306**

A DNA sequence <SEQ ID 10847> was identified in GBS which encodes amino acid sequence <SEQ ID 10848>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3307

- 15 A DNA sequence <SEQ ID 10849> was identified in GBS which encodes amino acid sequence <SEQ ID 10850>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3308

A DNA sequence <SEQ ID 10851> was identified in GBS which encodes amino acid sequence <SEQ ID 10852>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3309

- 20 A DNA sequence <SEQ ID 10853> was identified in GBS which encodes amino acid sequence <SEQ ID 10854>. Related sequences are <SEQ ID 10855>, <SEQ ID 10856>, <SEQ ID 10857>, <SEQ ID 10858>, <SEQ ID 10859>, <SEQ ID 10860>, <SEQ ID 10861>, <SEQ ID 10862>, <SEQ ID 10863>, <SEQ ID 10864>, <SEQ ID 10865> and <SEQ ID 10866>. These proteins and their epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3310**

A DNA sequence <SEQ ID 10867> was identified in GBS which encodes amino acid sequence <SEQ ID 10868>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3311

- 30 A DNA sequence <SEQ ID 10869> was identified in GBS which encodes amino acid sequence <SEQ ID 10870>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3312

A DNA sequence <SEQ ID 10871> was identified in GBS which encodes amino acid sequence <SEQ ID 10872>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3313

- 35 A DNA sequence <SEQ ID 10873> was identified in GBS which encodes amino acid sequence <SEQ ID 10874>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3314

A DNA sequence <SEQ ID 10875> was identified in GBS which encodes amino acid sequence <SEQ ID 10876>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3315

- 5 A DNA sequence <SEQ ID 10877> was identified in GBS which encodes amino acid sequence <SEQ ID 10878>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3316

A DNA sequence <SEQ ID 10879> was identified in GBS which encodes amino acid sequence <SEQ ID 10880>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 **Example 3317**

A DNA sequence <SEQ ID 10881> was identified in GBS which encodes amino acid sequence <SEQ ID 10882>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3318

- 15 A DNA sequence <SEQ ID 10883> was identified in GBS which encodes amino acid sequence <SEQ ID 10884>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3319

A DNA sequence <SEQ ID 10885> was identified in GBS which encodes amino acid sequence <SEQ ID 10886>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3320

- 20 A DNA sequence <SEQ ID 10887> was identified in GBS which encodes amino acid sequence <SEQ ID 10888>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3321

A DNA sequence <SEQ ID 10889> was identified in GBS which encodes amino acid sequence <SEQ ID 10890>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

25 **Example 3322**

A DNA sequence <SEQ ID 10891> was identified in GBS which encodes amino acid sequence <SEQ ID 10892>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3323

- 30 A DNA sequence <SEQ ID 10893> was identified in GBS which encodes amino acid sequence <SEQ ID 10894>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3324

A DNA sequence <SEQ ID 10895> was identified in GBS which encodes amino acid sequence <SEQ ID 10896>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3325

- 35 A DNA sequence <SEQ ID 10897> was identified in GBS which encodes amino acid sequence <SEQ ID 10898>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3326

A DNA sequence <SEQ ID 10899> was identified in GBS which encodes amino acid sequence <SEQ ID 10900>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3327

- 5 A DNA sequence <SEQ ID 10901> was identified in GBS which encodes amino acid sequence <SEQ ID 10902>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

Example 3328

A DNA sequence <SEQ ID 10903> was identified in GBS which encodes amino acid sequence <SEQ ID 10904>. This protein and its epitopes could be useful antigens for vaccines and/or diagnostics.

10 Example 3329

Seven rRNA genes were identified in *S. agalactiae*. These are SEQ IDs 12018 to 12024. These rRNA genes are particularly useful for diagnostic purposes and for phylogenetic studies. An alignment of the rRNA sequences is shown below:

```

15 12023 -----TTTCGAGTCAAAGTCATCAGCGTT
    12024 -----
    12019 -----TCCAATCATACTTAATTTCACTAATATCTGGATTTTGACATATTAGTTAATTCT
    12021 ...ATCGAATTGAACGGACTCAATTTGGTTGTTATGTAATTTT--ACATAATCTATGATTTCT
    12020 -----
    12018 -----
20 12022 -----CTTCCTTGTTTTCTTTAGAGATATTAACGTGA

    12023 TACTGTTACGGCAGCAGTTCCAAGAGTTACTCCACTCACAAGGACTGCTGATAATATTCT
    12024 -----
25 12019 TTTTCATGCTTTTGTAGATAAGCTACTTGTCTTTTTTTATTACTTTTTTACCTTTCTTT
    12021 TGCTCATGCTCTTTGAGATAGGCTAATTGTTCTTTTTTGTCATTTTTTATCTTTCTTC
    12020 -----
    12018 -----
    12022 CCCACTTTGGGCGTTAAATACTAAAGTAGCCTTTATTAAAGTTGATTTAGCAGCCCCA

30 12023 TTTTTTCATTTTTATTAACTACTCCTTTAC--GATAAGACATTAAATATTTTACCAAAA
    12024 -----
    12019 ACTGCTGACTGTTTGCTATTTTTTACTTCGTTTGACTGACTTTTAGATTCACTATTTCATT
    12021 ACTTCTGATTGCTTGCTATTTTTTACTTCGTTTGACTGAATTTTATGTTCACTATTTCATT
    12020 -----
35 12018 -----CTTT-GATACAATATTATCAAAATTATATTAA
    12022 TTTTCACCTGTTAAGGTAACAACTCCCACT-GTCTAAATGGTAATTAAACCCCTTCCAG

    12023 AATTCACGAAATATATTACGTCATTGTTACATTTATATTTGAAATCAACTATTTCTAAA
40 12024 -----
    12019 TGACAGCCTGCTAGTAACATCCCAATAATAGATATGGGAATTAACCATTTTACATATTTT
    12021 TGACAGCCTCCAAGTATCATCCCAAAATTGATATGGGAATTAACCATTTTATATATTTT
    12020 -----
    12018 CGGTAAAGATATGTTAAAGACCAAACTTGGATTATCAATCGT---TATCAAGAAATTA
45 12022 CA-CAGGATCGCTATCGTACTGAAAAGTAAGACCACTAAGTATATATATCGCATGATTA

    12023 TGAACCATAATCAAATCTAGAAAACGATAACCTTCTTCTATTCACTCT--ATCAATATA
    12024 -----
    12019 TTCAACATGCTCTCTTTTCTTAGAAAAATAAACTTCCCATGTCAGTATCTAATAAAAAATA
    12021 CTCATCATGTTCTCTTTTCTTAGAATAATAAATTTTATATATCAAGTATATAATGAAATTA
50 12020 -----
    12018 TTAGTG---ATTGTCTTTAGGAAGCACTA-----TTCAGAGA---AATTACTCG
    12022 CCTTCT--AATTCTCTAGAGAAAAGATCAAGAAAACGTTCTAAAACG--ACCTTTTCG

    12023 ATTACTCCATAGTGAAACTAAAAGAGAAATAAAAAAGAGTATAATTACTCTTAAATTA
55 12024 -----
    12019 ATTATTATTTACCAGTATGTTAAACTAATATTAGTATAACAAA-TTTTCACGAGTTTAA

```

-2954-

12021 ACTATTATTACCAACATTATATAAAATTAATTTTAGTATAACAAAATTTTCACGTATTTT
12020 -----ATCAAAAAACATGACCAGTATGAATTAAAGCAACGTATAATCAATGCCT
12018 CTCTAT--AGAGCAGCTAGCTTCACTTCCCATAGAAAATAATCAGTTTAT--ATGAT--
12022 TCCTTTGAAAAATGATTTACTAATCTTCCGTAAACCCCTAACGTATTTGTCATGATGATGT

5

12023 TAATATTTACGGAGAATAAGGGATTGGAACCCCTGCGCCAGTTACCCGACCTAACGATTT
12024 -----
12019 TT--TTTTAGTCGTAACATATACACTGAAAAATCTTATTTATTTTATACTACCTATCTATC
12021 ATAGTTTGTAGTCTTAACATGTAAACAGAAA-----A-----TC
10 12020 TAATGCGTAAAGGATACCAAGTACGAAGATA-----TC
12018 ---TGTTTTTTAGCAGCCGGTGAAGATA-----ACAACGCAAGTT
12022 GTGTGTTTCATCTGCAATGGGTTTAGCAAGT---TCA-----GATAACTCAAAATA

15

12023 AGCAAAACCGTCCTCTTCAGCCTCTTGAG--TAATCTCCAATTAATATTAATGGGCACG
12024 -----
12019 ATTCAAAACACTTTTATTACTTCAGAACCTATGACATTTAGGAGTCCCTTTGAATTC
12021 ATTGTGATA-----T-----TTTAAATGCCCTAATTAAATT--
12020 AAAAGTGC-----T-----TTAAGAGAATATTTATAAGAT--
12018 AGTTGCA--ACGTTTTTAAATCAAAATGA--CATTCCTGCAAGATATGTTTCATCCAAACGA
20 12022 AGTAATACGAGCATCTTTAGAATCTTTA--TTCGCTTCAACATATCCTGAGA--AATTAA

25

12023 AGTGGACTCGAACCACCGACCTCAGCCTTATCAGGCGTGCGCTCTAACCACTGAGCTAC
12024 -----
12019 ATTTAAATGTTGAGTCTCCACTAATCTTTGAAAAATTTCTTATTTCTGCTTGTTTT
12021 -----AATAATT-----AATATTATTATATATATA
12020 -----AATAACTCTCAGACGATGATT--TTACAGA
12018 AGCAGGAATTATTGTAATAAGAACCATG--TAATGCACGAATTATT--CCAG--GA
12022 ACTTTTTACTGCTTTAGTTACAGCTGCCTGACTAATATTTAACTTCTTAGCTAAATCAGA

30

12023 GCGCCCAAGCAAATGCTTGGTTTTACTTTTATGTAAAGTAAGCGGGTGACGAGAATCGA--
12024 -----
12019 AAACCTTCTATAACGATTGCAATAATGAAAAACAAATATAAGTAATTTTCAGTAACTTTT
12021 AATTCCTTCTACAATGA-----AAAAATAAATATAT--A--TTACAAGTAACATT--
12020 AAAT---TATGATAA---A---CTATAACAGACGTAT--AAATTGTAGAAAGTTG--
35 12018 AGTTATGATAAGATTGA-----GAACCTATGTCTATACAATGAGGTTCTTGTATCCCT
12022 ATTTGTCAACTGCTCTT-----GTGATAAAAGCATCAGAATGTGTTCTTGCGTATTAGT

40

12023 -ACTCGCGACAACAGCTTGGGAAGCTGTAGTTTTACCACTAACTACACCCGCTAAAAAC
12024 -----
12019 TCTCAAAATTACAGCACAAATCAAAAAAGACAGGCTTCTAAACCTTGTCTTTATAAAT
12021 --TCACAATAAATATCTAGTAGAAAAAGACAAGGTTTAGAAACCTTGTCTTTATAAGT
12020 ---GTAGGCTATGAGATTACCTAAAGAAGGCGACTTTATTACAATTCAAAGTTACAAC
12018 GGATTT-----TTTGG--AGTCACAGAAGATAAC--CAAATTTGTACCTTTTCAAGA
45 12022 CAATTTAA--CATCACTTTGACAAGTACCACAAATAATTCATGTTGATTTTCTGCTTTAA

50

12023 TTATATAATAAATGGCGCGAGACGGAATCGAACCGCCGACACATGGAGCTTCAATCCATT
12024 -----
12019 ATACCGGCGGCGGGGTGCAACCGGCACGTCCGTGAGGACACTGGATTTTGTAGTCCAGCG
12021 ATACCGGCGGCGGGGTGCAACCGGCACGTCCGTGAGGACACTGGATTTTGTAGTCCAGCG
12020 ATGATGGTAGTTTACACCGAATCTG-----GCGTGACACCA--TGATATTAAAAACAACCG
12018 GGGGGATCTGACATTACTGGATC-----CCTAATTGC-----AGCAGGCATAAA
12022 GCAAGATTTGAC--TCACTAAATGG-----TCTAATTTTGTCTTAAACTGTCTATATA

55

12023 GCTCTACCAACTGAGCTACCGAGCCTATTGCGGGAGCAGGATTTGAACCTACGACCTTCG
12024 -----
12019 CGTCTGCCAATTCGCCACGCGGCTATCTTAAACTGGGGTAGCTGGATTTCGA--ACCA
12021 CGTCTGCCAATTCGCCACGCGGCTATCTTAAACTGGGGTAGCTGGATTTCGA--ACCA
12020 AAAATGCC--CTCATTGGTGTTAATGATCAT--ACTTTAGTAACAGAAAATGATGGTCG
12018 AGCAGACCT--TTATGAGAACTTCACAGATGT---TGATGGTATATTTGCAGCACATCCA
60 12022 TACCT--CTT--TTTTGTAAACAGTAAATATATCACGAAGATATAGAAGAATCAATCATA

65

12023 GGTTA--TGAGCCCGACGAGCTACCTAGCTGCTCCA-----TCCCGCGATATCTTTAA
12024 -----
12019 ACGCA--TGAGGGAGTCAAAAGTCCCTTGCCTTACCG-----CTTGGCTATACCCCATGA
12021 ACGCA--TGAGGGAGTCAAAAGTCCCTTGCCTTACCG-----CTTGGCTATACCCCATGA
12020 ACGC--TGGGTGACACGAGAGCC--TGCAATA-----GTATACCTTTCATA
12018 GGT-----GTAGTTAAGAACCTCAGCTA-----TCCCTGAGCTTACTTATA
12022 GATAGGTGAAGAAGATAAAACCTTTTATCTCAACAACCTAATTTATAAACTTCTTTGCA

-2955-

5	12023	GGA-----GGATGTGGGATTTCGAACCCACGCACGCTTTTACAC--GCCTGACGGTT
	12024	-----
	12019	AAAGGCG----AGTGATGGGAATCGAACCCACGAATGTCAGAGCCACAATCTGATGTGT
	12021	AAAGGCG----AGTGATGGGAATCGAACCCACGAATGTCAGAGCCACAATCTGATGTGT
	12020	AAA-----AATACTGG---T-----TT--AACATTATCGCTA----TGATACGT
10	12018	AAGA-----AATGCGTGAAATAGCCTATGCGGGTTTTCGGTTT-TACATGATGAA-
	12022	AAAACCTTTCATACTATTAAAAACACGATCAGCTTTTTTCTCTGTAG-AACACATTGAAA
	12023	TTCAAGACCGTTCCTTTCAGCCGACTTGGGTAATCCTCCATATAACAAAAAATATGGAC
	12024	-----
	12019	TAACCACTTCACCACACCCGCCATATTAGAAAAAACACGGGCAGTAGGAATCGAACCCAC
15	12021	TAACCACTTCACCACACCCGCCATATTAGAAAAAACACGGGCAGTAGGAATCGAACCCAC
	12020	GAAACTGGTGTCTCCTACTATTGTAAATCTAGCAAGT-----CCGTATATCTTGGACCC--
	12018	-----GCTTTACTTCTGCCTATCGTGGCAGAATCCCTCTTGTATTAAAAATAC--
	12022	AAACAGTTGGTCCACTTCTGTGC-ATTAATGCAACATCGGCTCCAGAAATTAACATAC--
	12023	CTTGTAGGACTCGAACCTACGACCGCTCGGTTATGAGCCGAGTGCTCTAACAGTTGAGC
20	12024	-----
	12019	ACTGAAGGTTTTTGGAGACCTTAGTTCTACCTTTAAACTATGCCGTTTTACTATGGAGAGA
	12021	ACTGAAGGTTTTTGGAGACCTTAGTTCTACCTTTAAACTATGCCGTTTTACTATGGAGAGA
	12020	--TGAAGCACTCAAGTATATTGACTATGACCTTGATGTCAAAGTATTTGCAGATGGTGAA
	12018	----AAA---TAATCCCAACAGCCTGGTACAAAAATAGTTTTTAAAGCATACTCGTAG-
25	12022	----GTTCTTTTATGTACTTTATACTGGATTTTITAGTAATTGTAATATCCTCGAGTGAA
	12023	TAAAGGTCCAAAGTCTCAATAAAATAAATAGCGGCGGAGGGGATCGAACCCCGACCTCC
	12024	-----
	12019	GAGGGATTTCGAACCCCCGAACCCGAAGGAGCGGATTTACAGTCCGCCGCGTTTTAGCCTCT
	12021	GAGGGATTTCGAACCCCCGAACCCGAAGGAGCGGATTTACAGTCCGCCGCGTTTTAGCCTCT
30	12020	AAAAGACTACTAGATGTGGACGAATATGAACAGCATAAAGYTCAGATGAACT--ATCCTA
	12018	--TAACATAGCAGTAACTGG-GATCGCT--TCTGATAGCGTTTTTGCTAGCATAAACGTA
	12022	TTTCCCATAGATTTGACCATTAACTGATAATCTGATGACAAAATAGCAGACTTTAATAAA
	12023	CGGGTATG-AAACCGGACGCTCTAGCCAGCT--GAGCTACACCGCCATAAAAAATATATCCA
	12024	-----
35	12019	TCGCTATC-TCTCCTAAGGTATAAATGGCGCGAGACGGAATCGAACCGCCGACACATGGA
	12021	TCGCTATC-TCTCCTAAGGTATAAATGGCGCGAGACGGAATCGAACCGCCGACACATGGA
	12020	CCGATATT-GATTATATATTAAAGGAAAATGTAAAAATATTGGTAGAATGGATAAATGAG
	12018	TCTAAAT--ACTTAATGAATAGA--GAAGTAGGTTTCGGCCGAAAAG---TACTACAA
	12022	TCAATATCAACTCTACTTATAGACTTACCAATCAATATCTCTAAAAATGGATTTAGTTGAA
40	12023	TCGGGAAGACAGGATTTCGAACCTGCGACACCTTGGTCCCAAACCAAGTACTCTACCAAGC
	12024	-----
	12019	GCTTCAATCCATTGCTCTACCAACTGAGCTACCGAGCCTATTGCGGGAGCAGGATTTGAA
	12021	GCTTCAATCCATTGCTCTACCAACTGAGCTACCGAGCCTATTGCGGGAGCAGGATTTGAA
	12020	AATAAAGGCCCTTTTC-ATCATC--ATATATCAA-TATCTGGTATAAACGGTA-----
45	12018	ATTTTAGAG--GATTTAAATATT--AGTTTGAACATATGCCAACTGGCATAGATGAT
	12022	ATACCAAAATCCGGCTTAACCAGA--ACTATCCAACATGGTCTCAATGTGCGTAAGGGT
	12023	TCGGGAGACAGGATTTCGAACCTGCGACACCTTGGTCCCAAACCAAGTACTCTACCAAGC
	12024	-----
	12019	GCTTCAATCCATTGCTCTACCAACTGAGCTACCGAGCCTATTGCGGGAGCAGGATTTGAA
50	12021	GCTTCAATCCATTGCTCTACCAACTGAGCTACCGAGCCTATTGCGGGAGCAGGATTTGAA
	12020	AATAAAGGCCCTTTTC-ATCATC--ATATATCAA-TATCTGGTATAAACGGTA-----
	12018	ATTTTAGAG--GATTTAAATATT--AGTTTGAACATATGCCAACTGGCATAGATGAT
	12022	ATACCAAAATCCGGCTTAACCAGA--ACTATCCAACATGGTCTCAATGTGCGTAAGGGT
	12023	TGAGCTACTTCCCGAAAAATATGCAC--CCTAGAGGAGTCGAACCTCTAACCGCCTGATT
55	12024	-----
	12019	CCTACGACCTTCGGGTTATGAGCCCG--ACGAGCTACCTAGCTGCTCCATCCCGCGATAT
	12021	CCTACGACCTTCGGGTTATGAGCCCG--ACGAGCTACCTAGCTGCTCCATCCCGCGATAT
	12020	-----CCTTGAATTGAAA-----AAGCGCTAACCTAAC-ACACTAAATAGTG-TGT
	12018	CTATCCATTGT---CTTACGTGAAA--AAGAATTGACACCAATCAAAGAACAGAAATC
60	12022	TTAACAATTTACCTTTACCTAATACTAACGAACATCCCCACCAAGACAATAAGGAACA
	12023	CGTAGTCAG---GTACTCTATCCAGTTGAGCTAAGGGTGCTAAATATTATA-----TGCC
	12024	-----
	12019	CTTTAAAGGAGGATGTGGGATTTCGAACCCACGCACGCTTTTACACGCCTGACG--GTTTT
	12021	CTTTAAAGGAGGATGTGGGATTTCGAACCCACGCACGCTTTTACACGCCTGACG--GTTTT
65	12020	TTTTATTA---ATATCAAATTTAATTACA--ATACTATTGCAAAAATAT---ATACT
	12018	TTAAATTACCTAATCGTAACTAGAAGTAG--ATTACGTTGACATCCAA-----
	12022	TC--ACTACC-AATTTTAAACCAATAGCAACCATTCGTCATAGTCCATTTGAAGATTCT
	12023	GAGGACCGGAATC---GAACCGGTACGATGTTTTACCATCGCAGGATTTTAAAGTCTGTG
	12024	-----
	12019	CAAGACCGTTCCCTTCAGCCGGAATTGGGTAATCCTCCATATAACAAAAAATAGTCCGTA

	12021	CAAGACCGTTCCCTTCAGCCGGACTTGGGTAAATCCTCCATATAACA AAAAATAGTCCGTA
	12020	TAAAATAAA-----AAAAGTAGAAGATCAC TTTCTACTTTTTTAAAGAATAGTCCGTA
	12018	CACAATCTATC-----TACAATCGTAATGTAGGTGAAA-ATATGAAAAGTCAGATTG
5	12022	CATAATCGATT-----AAGAGCTCTTATGTAGCAGCAGCATCAGTAGAACCCACCCC
	12023	CGTCTGCCAGTTCGCCACCCCGGCCTCTAACAAGCGAACGACGGGGTT CGAACCCGCGA
	12024	-----
	12019	CGGGATT CGAACCCCGTGTTACC GCCGTGAAAAGGCGGTGTCTTAA CCCCCTTGACCAACGG
	12021	CGGGATT CGAACCCCGTGTTACC GCCGTGAAAAGGCGGTGTCTTAA CCCCCTTGACCAACGG
10	12020	CGGGATT CGAACCCCGTGTTACC GCCGTGAAAAGGCGGTGTCTTAA CCCCCTTGACCAACGG
	12018	GAGTCACTGCAACAGCGACACAAGCCTTATC-----AAGAGAAAAA-----ATCAATAT
	12022	CAGTC-CTGCACAGACAGGAATGGATT TTCTAATCTAATATGAACACCTTTATTAATAC
	12023	-CCCTCACCTT-----GGCAAGGTGATGTTCTACCACTGAACTACGTT CGCACTAAAGAC
15	12024	-----
	12019	-ACCATATTCTTGATGGGCACGAGTGGACTCGAACCACCGACCTCAGCTTATCAGGCGT
	12021	-ACCATATTCTTGATGGGCACGAGTGGACTCGAACCACCGACCTCAGCTTATCAGGCGT
	12020	-ACCATATTCTTGATGGGCACGAGTGGACTCGAACCACCGACCTCAGCTTATCAGGCGT
20	12018	CACCATGAT----ATCACA----AGGTTCAAGCGAA--GTCTCCATTATGT-----
	12022	CATATTGATTTTTGATTATATCTGCAGCTTTAAACACATCATTATCATTATTTAAAGGCA
	12023	ACTATTTATCCTATAAAAATTGTAATGCCGGC-----
	12024	-----TATCCTATAAAAATTGTAATGCCGGC-----
25	12019	GCGCTCTAACCACTGAGCTACGCGCCCCAAAATAACTTCTAAAATTATAAAGTTAATGCC
	12021	GCGCTCTAACCACTGAGCTACGCGCCCCAAG-----CTA-----
	12020	GCGCTCTAACCACTGAGCTACGCGCCCCAAG-----CTA-----
	12018	--TCGTTATAAACAGTAAGGATGAAAAAGAG-----CTAT
	12022	TTTGTCTACTATCAGAACTCGATAACAATACAAT----CTT-----CCTT
		* * *
30		
	12023	---TACATGACTTGAACACGCGACCCCTCTGATTACAAATCAGATGCTCTACCAACTGAGC
	12024	---TACATGACTTGAACACGCGACCCCTCTGATTACAAATCAGATGCTCTACCAACTGAGC
	12019	GGCTACATGACTTGAACACGCGACCCCTCTGATTACAAATCAGATGCTCTACCAACTGAGC
35	12021	--TTGCTTGGTTT-----T--TACTTTCTATA-----A-----
	12020	--TTGCTTGGTTT-----T--TACTTTCTATA-----A-----
	12018	TAAAGCACTATATGAA-ACAT--TCTTCCAA--AATAGTACCTATTACACTACTTACAC
	12022	TAGCTCAGAAATGGTA-ACGTAGTCATTAAAGTCAATACTAACCAT AATCATAGCTAATT
		* *
40		
	12023	TAAGCCGGCAATCTACTAATGCGGGTGAAGGGACTTGAACCCCCACGCGTTAAGCGCCA
	12024	TAAGCCGGCAATCTACTAATGCGGGTGAAGGGACTTGAACCCCCACGCGTTAAGCGCCA
	12019	TAAGCCGGCAATCTACTAATGCGGGTGAAGGGACTTGAACCCCCACGCGTTAAGCGCCA
	12021	-----AG-----TAAAGCGGGTGACGAGAATCGAACTC-----
45	12020	-----AG-----TAAAGCGGGTGACGAGAATCGAACTC-----
	12018	TATTAGATAGATAA--CAAATCGTCCT----AAGTAAGCTTA-----CTTAGGACGA
	12022	CATGATAAACCATCGT-CACATCGTCCTTTAATATCTAATCCTAAATTAAGTTTGGCAGGA
		* ** *
	12023	GATCCTAAATCTGGTGCGTCTGCCAATTCGCCCACACCCGCATTTCTAAATGACCCGTAC
50	12024	GATCCTAAATCTGGTGCGTCTGCCAATTCGCCCACACCCGCATTTCTAAATGACCCGTAC
	12019	GATCCTAAATCTGGTGCGTCTGCCAATTCGCCCACACCCGCATTTCTAAATGACCCGTAC
	12021	-----GCGACAACAGC-----
	12020	-----GCGACAACAGC-----
55	12018	TTTT----ATTTAGAACATAGGATAGTTTTCCACTTTTAAATCGTAA-----CCACTT
	12022	GCTT----TCTCAAAATTTTCATAAAACCTCCC TAATAAAATATAGAA-T-ATCCATAT
		*
	12023	TGGGCTCGAACCAGTGACCCATTGATTAAAAGTCAATTGCTCTACCAACTGAGCTAACGA
60	12024	TGGGCTCGAACCAGTGACCCATTGATTAAAAGTCAATTGCTCTACCAACTGAGCTAACGA
	12019	TGGGCTCGAACCAGTGACCCATTGATTAAAAGTCAATTGCTCTACCAACTGAGCTAACGA
	12021	-----T-----TGGAAGGCTGTAGTTTTTAACTA-----
	12020	-----TGGAAGGCTGTAGTTTTTAACTA-----
	12018	GGTATCA-----GTGACA---AATTCGGA--CAATTAAGATGTTAGCCAATCTTAAGG
65	12022	TATAACATAACAAATGACA---AATTCGGA--CAATTAAGATGCTAGCCAATCTTAAGG
		* * * * *
	12023	GTCTACGGTCCCGACGGGAATCGAACC CGCATCTTCGCCGTGACAGGGCGACGTGATAA
	12024	GTCTACGGTCCCGACGGGAATCGAACC CGCATCTTCGCCGTGACAGGGCGACGTGATAA

-2957-

12019 GTCTACGGTCCCGACGGGAATCGAACCCGCGATCTTCGCCGTGACAGGGCGACGTGATAA
12021 -----CACC-----
12020 -----CACC-----
5 12018 ATA-ATAATTCCAATAAAAA-----AAGGCTAACCAAAGTTAGTC
12022 ATA-ATAATTCCAATAAAAA-----AAGGCTAACCAAAGTTAGTC
*

12023 CCGCTACACTACGGGACCTATGGGAGTTAACGGGATCGAACCGCTGACCCCTCTGCTTGTA
12024 CCGCTACACTACGGGACCTATGGGAGTTAACGGGATCGAACCGCTGACCCCTCTGCTTGTA
10 12019 CCGCTACACTACGGGACCTATGGGAGTTAACGGGATCGAACCGCTGACCCCTCTGCTTGTA
12021 -CGCT-----TCTATGGGAGTTAACGGGATCGAACCGCTGACCCCTCTGCTTGTA
12020 -CGCT-----TCTATGGGAGTTAACGGGATCGAACCGCTGACCCCTCTGCTTGTA
12018 TCCCTTTA-----TCTACTCCGCCAGTAGGACTCGAACCTACGACATCATGATTAAAC
12022 TCCCTTTA-----TCTACTCCGCCAGTAGGACTCGAACCTACGACATCATGATTAAAC
15 * ** *** ** ***** ** **

12023 AGGCAGATGCT-CTCCCAGCTGAGCTAAACTCCCTTT--GCTAAGCGACTACCTTATCTC
12024 AGGCAGATGCT-CTCCCAGCTGAGCTAAACTCCCTTT--GCTAAGCGACTACCTTATCTC
20 12019 AGGCAGATGCT-CTCCCAGCTGAGCTAAACTCCCTTT--GCTAAGCGACTACCTTATCTC
12021 AGGCAGATGCT-CTCCCAGCTGAGCTAAACTCCCTTT--GCTAAGCGACTACCTTATCTC
12020 AGGCAGATGCT-CTCCCAGCTGAGCTAAACTCCCTTT--GCTAAGCGACTACCTTATCTC
12018 AGTCATGCGCTACTACCAACTGAGCTATGGCGGATTATAGCTAAGCGACTACCTTATCTC
12022 AGTCATGCGCTACTACCAACTGAGCTATGGCGGATTATAGCTAAGCGACTACCTTATCTC
25 ** ** *** ** ** ***** ** *****

12023 ACAGGGGGCAACCCCAACTACTTCCGGCGTTCTAGGGCTTAACCTTCTGTGTTCCGGCATG
12024 ACAGGGGGCAACCCCAACTACTTCCGGCGTTCTAGGGCTTAACCTTCTGTGTTCCGGCATG
30 12019 ACAGGGGGCAACCCCAACTACTTCCGGCGTTCTAGGGCTTAACCTTCTGTGTTCCGGCATG
12021 ACAGGGGGCAACCCCAACTACTTCCGGCGTTCTAGGGCTTAACCTTCTGTGTTCCGGCATG
12020 ACAGGGGGCAACCCCAACTACTTCCGGCGTTCTAGGGCTTAACCTTCTGTGTTCCGGCATG
12018 ACAGGGGGCAACCCCAACTACTTCCGGCGTTCTAGGGCTTAACCTTCTGTGTTCCGGCATG
12022 ACAGGGGGCAACCCCAACTACTTCCGGCGTTCTAGGGCTTAACCTTCTGTGTTCCGGCATG

12023 AGAACAGGTGTATCTCCTAGGCAATTATCACTTAACTATTGAGCCTTATTCACCTCAAAT
12024 AGAACAGGTGTATCTCCTAGGCAATTATCACTTAACTATTGAGCCTTATTCACCTCAAAT
35 12019 AGAACAGGTGTATCTCCTAGGCAATTATCACTTAACTATTGAGCCTTATTCACCTCAAAT
12021 AGAACAGGTGTATCTCCTAGGCAATTATCACTTAACTATTGAGCCTTATTCACCTCAAAT
12020 AGAACAGGTGTATCTCCTAGGCAATTATCACTTAACTATTGAGCCTTATTCACCTCAAAT
40 12018 AGAACAGGTGTATCTCCTAGGCAATTATCACTTAACTATTGAGCCTTATTCACCTCAAAT
12022 AGAACAGGTGTATCTCCTAGGCAATTATCACTTAACTATTGAGCCTTATTCACCTCAAAT

12023 TGAATATCTATAGTCTAACAAGAAACCGTAACGTTGTCAATATCTCTTTTGGATAAGTC
12024 TGAATATCTATAGTCTAACAAGAAACCGTAACGTTGTCAATATCTCTTTTGGATAAGTC
45 12019 TGAATATCTATAGTCTAACAAGAAACCGTAACGTTGTCAATATCTCTTTTGGATAAGTC
12021 TGAATATCTATAGTCTAACAAGAAACCGTAACGTTGTCAATATCTCTTTTGGATAAGTC
12020 TGAATATCTATAGTCTAACAAGAAACCGTAACGTTGTCAATATCTCTTTTGGATAAGTC
50 12018 TGAATATCTATAGTCTAACAAGAAACCGTAACGTTGTCAATATCTCTTTTGGATAAGTC
12022 TGAATATCTATAGTCTAACAAGAAACCGTAACGTTGTCAATATCTCTTTTGGATAAGTC

12023 CTCGAGCTATTAGTATTAGTCCGCTAAATGTGTCAACCAATTACACTCCTAACCTATCT
12024 CTCGAGCTATTAGTATTAGTCCGCTAAATGTGTCAACCAATTACACTCCTAACCTATCT
55 12019 CTCGAGCTATTAGTATTAGTCCGCTAAATGTGTCAACCAATTACACTCCTAACCTATCT
12021 CTCGAGCTATTAGTATTAGTCCGCTAAATGTGTCAACCAATTACACTCCTAACCTATCT
12020 CTCGAGCTATTAGTATTAGTCCGCTAAATGTGTCAACCAATTACACTCCTAACCTATCT
12018 CTCGAGCTATTAGTATTAGTCCGCTAAATGTGTCAACCAATTACACTCCTAACCTATCT
60 12022 CTCGAGCTATTAGTATTAGTCCGCTAAATGTGTCAACCAATTACACTCCTAACCTATCT

12023 ACCTGATCATCTCTCAGGGCTCTTACTGATATAAAATCATGGGAAATCTCATCTTGAGGT
12024 ACCTGATCATCTCTCAGGGCTCTTACTGATATAAAATCATGGGAAATCTCATCTTGAGGT
65 12019 ACCTGATCATCTCTCAGGGCTCTTACTGATATAAAATCATGGGAAATCTCATCTTGAGGT
12021 ACCTGATCATCTCTCAGGGCTCTTACTGATATAAAATCATGGGAAATCTCATCTTGAGGT
12020 ACCTGATCATCTCTCAGGGCTCTTACTGATATAAAATCATGGGAAATCTCATCTTGAGGT
12018 ACCTGATCATCTCTCAGGGCTCTTACTGATATAAAATCATGGGAAATCTCATCTTGAGGT
12022 ACCTGATCATCTCTCAGGGCTCTTACTGATATAAAATCATGGGAAATCTCATCTTGAGGT

-2958-

5 12023 GGGCTTCGCACCTAGATGCTTTAGCGCTTATCCCTTCCCTACATAGCTACCCAGCGATG
12024 GGGCTTCGCACCTAGATGCTTTAGCGCTTATCCCTTCCCTACATAGCTACCCAGCGATG
12019 GGGCTTCGCACCTAGATGCTTTAGCGCTTATCCCTTCCCTACATAGCTACCCAGCGATG
12021 GGGCTTCGCACCTAGATGCTTTAGCGCTTATCCCTTCCCTACATAGCTACCCAGCGATG
12020 GGGCTTCGCACCTAGATGCTTTAGCGCTTATCCCTTCCCTACATAGCTACCCAGCGATG
12018 GGGCTTCGCACCTAGATGCTTTAGCGCTTATCCCTTCCCTACATAGCTACCCAGCGATG
10 12022 GGGCTTCGCACCTAGATGCTTTAGCGCTTATCCCTTCCCTACATAGCTACCCAGCGATG

15 12023 CCTTTGGCAAGACAACCTGGTACACCAGCGGTAAGTCCACTCTGGTCCTCTCGTACTAGGA
12024 CCTTTGGCAAGACAACCTGGTACACCAGCGGTAAGTCCACTCTGGTCCTCTCGTACTAGGA
12019 CCTTTGGCAAGACAACCTGGTACACCAGCGGTAAGTCCACTCTGGTCCTCTCGTACTAGGA
12021 CCTTTGGCAAGACAACCTGGTACACCAGCGGTAAGTCCACTCTGGTCCTCTCGTACTAGGA
12020 CCTTTGGCAAGACAACCTGGTACACCAGCGGTAAGTCCACTCTGGTCCTCTCGTACTAGGA
12018 CCTTTGGCAAGACAACCTGGTACACCAGCGGTAAGTCCACTCTGGTCCTCTCGTACTAGGA
12022 CCTTTGGCAAGACAACCTGGTACACCAGCGGTAAGTCCACTCTGGTCCTCTCGTACTAGGA

20 12023 GCAGATCCTCTCAAATTTCTACGCCCGCGACGGATAGGGACCGAACTGTCTCAGCAGCT
12024 GCAGATCCTCTCAAATTTCTACGCCCGCGACGGATAGGGACCGAACTGTCTCAGCAGCT
12019 GCAGATCCTCTCAAATTTCTACGCCCGCGACGGATAGGGACCGAACTGTCTCAGCAGCT
25 12021 GCAGATCCTCTCAAATTTCTACGCCCGCGACGGATAGGGACCGAACTGTCTCAGCAGCT
12020 GCAGATCCTCTCAAATTTCTACGCCCGCGACGGATAGGGACCGAACTGTCTCAGCAGCT
12018 GCAGATCCTCTCAAATTTCTACGCCCGCGACGGATAGGGACCGAACTGTCTCAGCAGCT
12022 GCAGATCCTCTCAAATTTCTACGCCCGCGACGGATAGGGACCGAACTGTCTCAGCAGCT

30 12023 TCTGAACCCAGCTCGCGTGCCGCTTTAATGGGCGAACAGCCCAACCTTGGGACCGACTA
12024 TCTGAACCCAGCTCGCGTGCCGCTTTAATGGGCGAACAGCCCAACCTTGGGACCGACTA
12019 TCTGAACCCAGCTCGCGTGCCGCTTTAATGGGCGAACAGCCCAACCTTGGGACCGACTA
35 12021 TCTGAACCCAGCTCGCGTGCCGCTTTAATGGGCGAACAGCCCAACCTTGGGACCGACTA
12020 TCTGAACCCAGCTCGCGTGCCGCTTTAATGGGCGAACAGCCCAACCTTGGGACCGACTA
12018 TCTGAACCCAGCTCGCGTGCCGCTTTAATGGGCGAACAGCCCAACCTTGGGACCGACTA
12022 TCTGAACCCAGCTCGCGTGCCGCTTTAATGGGCGAACAGCCCAACCTTGGGACCGACTA

40 12023 CAGCCCCAGGATGCGACGAGCCGACATCGAGGTGCCAAACCTCCCGTCGATGTGAATCTC
12024 CAGCCCCAGGATGCGACGAGCCGACATCGAGGTGCCAAACCTCCCGTCGATGTGAATCTC
12019 CAGCCCCAGGATGCGACGAGCCGACATCGAGGTGCCAAACCTCCCGTCGATGTGAATCTC
45 12021 CAGCCCCAGGATGCGACGAGCCGACATCGAGGTGCCAAACCTCCCGTCGATGTGAATCTC
12020 CAGCCCCAGGATGCGACGAGCCGACATCGAGGTGCCAAACCTCCCGTCGATGTGAATCTC
12018 CAGCCCCAGGATGCGACGAGCCGACATCGAGGTGCCAAACCTCCCGTCGATGTGAATCTC
12022 CAGCCCCAGGATGCGACGAGCCGACATCGAGGTGCCAAACCTCCCGTCGATGTGAATCTC

50 12023 TTGGGGGAGATAAGCCTGTTATCCCCAGGGTAGCTTTTATCCGTTGAGCGATGGCCCTTC
12024 TTGGGGGAGATAAGCCTGTTATCCCCAGGGTAGCTTTTATCCGTTGAGCGATGGCCCTTC
12019 TTGGGGGAGATAAGCCTGTTATCCCCAGGGTAGCTTTTATCCGTTGAGCGATGGCCCTTC
12021 TTGGGGGAGATAAGCCTGTTATCCCCAGGGTAGCTTTTATCCGTTGAGCGATGGCCCTTC
12020 TTGGGGGAGATAAGCCTGTTATCCCCAGGGTAGCTTTTATCCGTTGAGCGATGGCCCTTC
12018 TTGGGGGAGATAAGCCTGTTATCCCCAGGGTAGCTTTTATCCGTTGAGCGATGGCCCTTC
55 12022 TTGGGGGAGATAAGCCTGTTATCCCCAGGGTAGCTTTTATCCGTTGAGCGATGGCCCTTC

60 12023 CATACGGAACCACCGGATCACTAAGCCCGACTTTCGTCCCTGCTCGAGTTGTAGCTCTCG
12024 CATACGGAACCACCGGATCACTAAGCCCGACTTTCGTCCCTGCTCGAGTTGTAGCTCTCG
12019 CATACGGAACCACCGGATCACTAAGCCCGACTTTCGTCCCTGCTCGAGTTGTAGCTCTCG
12021 CATACGGAACCACCGGATCACTAAGCCCGACTTTCGTCCCTGCTCGAGTTGTAGCTCTCG
12020 CATACGGAACCACCGGATCACTAAGCCCGACTTTCGTCCCTGCTCGAGTTGTAGCTCTCG
12018 CATACGGAACCACCGGATCACTAAGCCCGACTTTCGTCCCTGCTCGAGTTGTAGCTCTCG
12022 CATACGGAACCACCGGATCACTAAGCCCGACTTTCGTCCCTGCTCGAGTTGTAGCTCTCG

65 12023 CAGTCAAGCTCCCTTATACCTTTTACACTCTACGACTGATTTCCAACCAAGTCTGAGGGAAC
12024 CAGTCAAGCTCCCTTATACCTTTTACACTCTACGACTGATTTCCAACCAAGTCTGAGGGAAC
12019 CAGTCAAGCTCCCTTATACCTTTTACACTCTACGACTGATTTCCAACCAAGTCTGAGGGAAC

-2959-

12021 CAGTCAAGCTCCCTTATACCTTTTACACTCTACGACTGATTTCCAACCAGTCTGAGGGAAC
12020 CAGTCAAGCTCCCTTATACCTTTTACACTCTACGACTGATTTCCAACCAGTCTGAGGGAAC
12018 CAGTCAAGCTCCCTTATACCTTTTACACTCTACGACTGATTTCCAACCAGTCTGAGGGAAC
12022 CAGTCAAGCTCCCTTATACCTTTTACACTCTACGACTGATTTCCAACCAGTCTGAGGGAAC

12023 CTTTGGGCGCCTCCGTTACCTTTTAGGAGGCGACCGCCCCAGTCAAAC TGCCCGTCAGAC
12024 CTTTGGGCGCCTCCGTTACCTTTTAGGAGGCGACCGCCCCAGTCAAAC TGCCCGTCAGAC
12019 CTTTGGGCGCCTCCGTTACCTTTTAGGAGGCGACCGCCCCAGTCAAAC TGCCCGTCAGAC
12021 CTTTGGGCGCCTCCGTTACCTTTTAGGAGGCGACCGCCCCAGTCAAAC TGCCCGTCAGAC
12020 CTTTGGGCGCCTCCGTTACCTTTTAGGAGGCGACCGCCCCAGTCAAAC TGCCCGTCAGAC
12018 CTTTGGGCGCCTCCGTTACCTTTTAGGAGGCGACCGCCCCAGTCAAAC TGCCCGTCAGAC
12022 CTTTGGGCGCCTCCGTTACCTTTTAGGAGGCGACCGCCCCAGTCAAAC TGCCCGTCAGAC

12023 ACTGTCTCCGATAGGGATTGCCTATCTGGGTTAGAGTAGCCATAACACAAGGGTAGTATC
12024 ACTGTCTCCGATAGGGATTGCCTATCTGGGTTAGAGTAGCCATAACACAAGGGTAGTATC
12019 ACTGTCTCCGATAGGGATTGCCTATCTGGGTTAGAGTAGCCATAACACAAGGGTAGTATC
12021 ACTGTCTCCGATAGGGATTGCCTATCTGGGTTAGAGTAGCCATAACACAAGGGTAGTATC
12020 ACTGTCTCCGATAGGGATTGCCTATCTGGGTTAGAGTAGCCATAACACAAGGGTAGTATC
12018 ACTGTCTCCGATAGGGATTGCCTATCTGGGTTAGAGTAGCCATAACACAAGGGTAGTATC
12022 ACTGTCTCCGATAGGGATTGCCTATCTGGGTTAGAGTAGCCATAACACAAGGGTAGTATC

12023 CCAACAACGCCTCAAACGAAACTGGCGTCCCGTTATCATAGGCTCCTACCTATCCTGTAC
12024 CCAACAACGCCTCAAACGAAACTGGCGTCCCGTTATCATAGGCTCCTACCTATCCTGTAC
12019 CCAACAACGCCTCAAACGAAACTGGCGTCCCGTTATCATAGGCTCCTACCTATCCTGTAC
12021 CCAACAACGCCTCAAACGAAACTGGCGTCCCGTTATCATAGGCTCCTACCTATCCTGTAC
12020 CCAACAACGCCTCAAACGAAACTGGCGTCCCGTTATCATAGGCTCCTACCTATCCTGTAC
12018 CCAACAACGCCTCAAACGAAACTGGCGTCCCGTTATCATAGGCTCCTACCTATCCTGTAC
12022 CCAACAACGCCTCAAACGAAACTGGCGTCCCGTTATCATAGGCTCCTACCTATCCTGTAC

12023 ATGTGGTACAGATACTCAATATCAAAC TGCAAGTAAAGCTCCATGGGGTCTTCCGTCCTG
12024 ATGTGGTACAGATACTCAATATCAAAC TGCAAGTAAAGCTCCATGGGGTCTTCCGTCCTG
12019 ATGTGGTACAGATACTCAATATCAAAC TGCAAGTAAAGCTCCATGGGGTCTTCCGTCCTG
12021 ATGTGGTACAGATACTCAATATCAAAC TGCAAGTAAAGCTCCATGGGGTCTTCCGTCCTG
12020 ATGTGGTACAGATACTCAATATCAAAC TGCAAGTAAAGCTCCATGGGGTCTTCCGTCCTG
12018 ATGTGGTACAGATACTCAATATCAAAC TGCAAGTAAAGCTCCATGGGGTCTTCCGTCCTG
12022 ATGTGGTACAGATACTCAATATCAAAC TGCAAGTAAAGCTCCATGGGGTCTTCCGTCCTG

12023 TCGCGGGTAACCTGCATCTTACAGGTAATAAAATTTACCGAGTCTCTCGTTGAGACAG
12024 TCGCGGGTAACCTGCATCTTACAGGTAATAAAATTTACCGAGTCTCTCGTTGAGACAG
12019 TCGCGGGTAACCTGCATCTTACAGGTAATAAAATTTACCGAGTCTCTCGTTGAGACAG
12021 TCGCGGGTAACCTGCATCTTACAGGTAATAAAATTTACCGAGTCTCTCGTTGAGACAG
12020 TCGCGGGTAACCTGCATCTTACAGGTAATAAAATTTACCGAGTCTCTCGTTGAGACAG
12018 TCGCGGGTAACCTGCATCTTACAGGTAATAAAATTTACCGAGTCTCTCGTTGAGACAG
12022 TCGCGGGTAACCTGCATCTTACAGGTAATAAAATTTACCGAGTCTCTCGTTGAGACAG

12023 TGCCCAAATCATTACGCCCTTTCGTGCGGGTCGGAAC TTACCCGACAAGGAATTTGCTAC
12024 TGCCCAAATCATTACGCCCTTTCGTGCGGGTCGGAAC TTACCCGACAAGGAATTTGCTAC
12019 TGCCCAAATCATTACGCCCTTTCGTGCGGGTCGGAAC TTACCCGACAAGGAATTTGCTAC
12021 TGCCCAAATCATTACGCCCTTTCGTGCGGGTCGGAAC TTACCCGACAAGGAATTTGCTAC
12020 TGCCCAAATCATTACGCCCTTTCGTGCGGGTCGGAAC TTACCCGACAAGGAATTTGCTAC
12018 TGCCCAAATCATTACGCCCTTTCGTGCGGGTCGGAAC TTACCCGACAAGGAATTTGCTAC
12022 TGCCCAAATCATTACGCCCTTTCGTGCGGGTCGGAAC TTACCCGACAAGGAATTTGCTAC

12023 CTTAGGACCGTTATAGTTACGGCCGCCGTTTACTGGGGCTTCAATTCATACCTTCGCTTA
12024 CTTAGGACCGTTATAGTTACGGCCGCCGTTTACTGGGGCTTCAATTCATACCTTCGCTTA
12019 CTTAGGACCGTTATAGTTACGGCCGCCGTTTACTGGGGCTTCAATTCATACCTTCGCTTA
12021 CTTAGGACCGTTATAGTTACGGCCGCCGTTTACTGGGGCTTCAATTCATACCTTCGCTTA
12020 CTTAGGACCGTTATAGTTACGGCCGCCGTTTACTGGGGCTTCAATTCATACCTTCGCTTA
12018 CTTAGGACCGTTATAGTTACGGCCGCCGTTTACTGGGGCTTCAATTCATACCTTCGCTTA
12022 CTTAGGACCGTTATAGTTACGGCCGCCGTTTACTGGGGCTTCAATTCATACCTTCGCTTA

-2960-

5	12023	CGCTAAGCACTCCTCTTAACCTTCCAGCACCGGGCAGGCGTCACCCCTATACATCATCT
	12024	CGCTAAGCACTCCTCTTAACCTTCCAGCACCGGGCAGGCGTCACCCCTATACATCATCT
	12019	CGCTAAGCACTCCTCTTAACCTTCCAGCACCGGGCAGGCGTCACCCCTATACATCATCT
	12021	CGCTAAGCACTCCTCTTAACCTTCCAGCACCGGGCAGGCGTCACCCCTATACATCATCT
	12020	CGCTAAGCACTCCTCTTAACCTTCCAGCACCGGGCAGGCGTCACCCCTATACATCATCT
10	12018	CGCTAAGCACTCCTCTTAACCTTCCAGCACCGGGCAGGCGTCACCCCTATACATCATCT
	12022	CGCTAAGCACTCCTCTTAACCTTCCAGCACCGGGCAGGCGTCACCCCTATACATCATCT

	12023	TACGATTTAGCAGAGAGCTGTGTTTTTGATAAACAGTTGCTTGGGCCATTCTACTGCGGC
	12024	TACGATTTAGCAGAGAGCTGTGTTTTTGATAAACAGTTGCTTGGGCCATTCTACTGCGGC
15	12019	TACGATTTAGCAGAGAGCTGTGTTTTTGATAAACAGTTGCTTGGGCCATTCTACTGCGGC
	12021	TACGATTTAGCAGAGAGCTGTGTTTTTGATAAACAGTTGCTTGGGCCATTCTACTGCGGC
	12020	TACGATTTAGCAGAGAGCTGTGTTTTTGATAAACAGTTGCTTGGGCCATTCTACTGCGGC
	12018	TACGATTTAGCAGAGAGCTGTGTTTTTGATAAACAGTTGCTTGGGCCATTCTACTGCGGC
	12022	TACGATTTAGCAGAGAGCTGTGTTTTTGATAAACAGTTGCTTGGGCCATTCTACTGCGGC
20		*****
	12023	TGATCTAAAATCAGCGCCCTTCTCCGAAGTTACGGGGCCATTTGCCGAGTTCCCTAA
	12024	TGATCTAAAATCAGCGCCCTTCTCCGAAGTTACGGGGCCATTTGCCGAGTTCCCTAA
	12019	TGATCTAAAATCAGCGCCCTTCTCCGAAGTTACGGGGCCATTTGCCGAGTTCCCTAA
	12021	TGATCTAAAATCAGCGCCCTTCTCCGAAGTTACGGGGCCATTTGCCGAGTTCCCTAA
25	12020	TGATCTAAAATCAGCGCCCTTCTCCGAAGTTACGGGGCCATTTGCCGAGTTCCCTAA
	12018	TGATCTAAAATCAGCGCCCTTCTCCGAAGTTACGGGGCCATTTGCCGAGTTCCCTAA
	12022	TGATCTAAAATCAGCGCCCTTCTCCGAAGTTACGGGGCCATTTGCCGAGTTCCCTAA

	12023	CGAGAGTTCTCTCGCTCACCTGAGGCTACTCGCCTCGACTACCTGTGTGCGTTTGC
30	12024	CGAGAGTTCTCTCGCTCACCTGAGGCTACTCGCCTCGACTACCTGTGTGCGTTTGC
	12019	CGAGAGTTCTCTCGCTCACCTGAGGCTACTCGCCTCGACTACCTGTGTGCGTTTGC
	12021	CGAGAGTTCTCTCGCTCACCTGAGGCTACTCGCCTCGACTACCTGTGTGCGTTTGC
	12020	CGAGAGTTCTCTCGCTCACCTGAGGCTACTCGCCTCGACTACCTGTGTGCGTTTGC
35	12018	CGAGAGTTCTCTCGCTCACCTGAGGCTACTCGCCTCGACTACCTGTGTGCGTTTGC
	12022	CGAGAGTTCTCTCGCTCACCTGAGGCTACTCGCCTCGACTACCTGTGTGCGTTTGC

40	12023	CGGGTAGAGTATATGTATCGCTAGAAGCTTTTCTTGGCAGTGTGACATCACTA
	12024	CGGGTAGAGTATATGTATCGCTAGAAGCTTTTCTTGGCAGTGTGACATCACTA
	12019	CGGGTAGAGTATATGTATCGCTAGAAGCTTTTCTTGGCAGTGTGACATCACTA
	12021	CGGGTAGAGTATATGTATCGCTAGAAGCTTTTCTTGGCAGTGTGACATCACTA
	12020	CGGGTAGAGTATATGTATCGCTAGAAGCTTTTCTTGGCAGTGTGACATCACTA
45	12018	CGGGTAGAGTATATGTATCGCTAGAAGCTTTTCTTGGCAGTGTGACATCACTA
	12022	CGGGTAGAGTATATGTATCGCTAGAAGCTTTTCTTGGCAGTGTGACATCACTA

50	12023	TACTAAACTTCGCTCCTCGTCACAGCTCAATGTTAAAGATATAAGCATTTGACTCATATC
	12024	TACTAAACTTCGCTCCTCGTCACAGCTCAATGTTAAAGATATAAGCATTTGACTCATATC
	12019	TACTAAACTTCGCTCCTCGTCACAGCTCAATGTTAAAGATATAAGCATTTGACTCATATC
	12021	TACTAAACTTCGCTCCTCGTCACAGCTCAATGTTAAAGATATAAGCATTTGACTCATATC
	12020	TACTAAACTTCGCTCCTCGTCACAGCTCAATGTTAAAGATATAAGCATTTGACTCATATC
55	12018	TACTAAACTTCGCTCCTCGTCACAGCTCAATGTTAAAGATATAAGCATTTGACTCATATC
	12022	TACTAAACTTCGCTCCTCGTCACAGCTCAATGTTAAAGATATAAGCATTTGACTCATATC

60	12023	ACACCTCACTGTTTGACCAGACACTTCCAATCGTCTGGTTTAGTTAGCCTACTGCGTCCC
	12024	ACACCTCACTGTTTGACCAGACACTTCCAATCGTCTGGTTTAGTTAGCCTACTGCGTCCC
	12019	ACACCTCACTGTTTGACCAGACACTTCCAATCGTCTGGTTTAGTTAGCCTACTGCGTCCC
	12021	ACACCTCACTGTTTGACCAGACACTTCCAATCGTCTGGTTTAGTTAGCCTACTGCGTCCC
	12020	ACACCTCACTGTTTGACCAGACACTTCCAATCGTCTGGTTTAGTTAGCCTACTGCGTCCC
65	12018	ACACCTCACTGTTTGACCAGACACTTCCAATCGTCTGGTTTAGTTAGCCTACTGCGTCCC
	12022	ACACCTCACTGTTTGACCAGACACTTCCAATCGTCTGGTTTAGTTAGCCTACTGCGTCCC

	12023	TCCATCACTATATACTCTAGTACAGGAATATCAACCTGTTGTCCATCGGATACACCTTTC
	12024	TCCATCACTATATACTCTAGTACAGGAATATCAACCTGTTGTCCATCGGATACACCTTTC
	12019	TCCATCACTATATACTCTAGTACAGGAATATCAACCTGTTGTCCATCGGATACACCTTTC
	12021	TCCATCACTATATACTCTAGTACAGGAATATCAACCTGTTGTCCATCGGATACACCTTTC

-2961-

	12020	TCCATCACTATATACTCTAGTACAGGAATATCAACCTGTTGTCCATCGGATACACCTTTC
	12018	TCCATCACTATATACTCTAGTACAGGAATATCAACCTGTTGTCCATCGGATACACCTTTC
	12022	TCCATCACTATATACTCTAGTACAGGAATATCAACCTGTTGTCCATCGGATACACCTTTC
5		*****
	12023	GGTCTCTCCTTAGGTCCCGACTAACCAGGGCGGACGAGCCTTCCCCTGGAACCTTAGT
	12024	GGTCTCTCCTTAGGTCCCGACTAACCAGGGCGGACGAGCCTTCCCCTGGAACCTTAGT
	12019	GGTCTCTCCTTAGGTCCCGACTAACCAGGGCGGACGAGCCTTCCCCTGGAACCTTAGT
	12021	GGTCTCTCCTTAGGTCCCGACTAACCAGGGCGGACGAGCCTTCCCCTGGAACCTTAGT
10	12020	GGTCTCTCCTTAGGTCCCGACTAACCAGGGCGGACGAGCCTTCCCCTGGAACCTTAGT
	12018	GGTCTCTCCTTAGGTCCCGACTAACCAGGGCGGACGAGCCTTCCCCTGGAACCTTAGT
	12022	GGTCTCTCCTTAGGTCCCGACTAACCAGGGCGGACGAGCCTTCCCCTGGAACCTTAGT

15	12023	CTTACGGTGGACAGGATTCTCACCTGTCTTGCGCTACTCATACCGGCATTCTCACTTCTA
	12024	CTTACGGTGGACAGGATTCTCACCTGTCTTGCGCTACTCATACCGGCATTCTCACTTCTA
	12019	CTTACGGTGGACAGGATTCTCACCTGTCTTGCGCTACTCATACCGGCATTCTCACTTCTA
	12021	CTTACGGTGGACAGGATTCTCACCTGTCTTGCGCTACTCATACCGGCATTCTCACTTCTA
	12020	CTTACGGTGGACAGGATTCTCACCTGTCTTGCGCTACTCATACCGGCATTCTCACTTCTA
20	12018	CTTACGGTGGACAGGATTCTCACCTGTCTTGCGCTACTCATACCGGCATTCTCACTTCTA
	12022	CTTACGGTGGACAGGATTCTCACCTGTCTTGCGCTACTCATACCGGCATTCTCACTTCTA

25	12023	TGCGTTCCAGCGCTCCTCACGGTACACCTTCTTACACATAGAACGCTCTCCTACCATGA
	12024	TGCGTTCCAGCGCTCCTCACGGTACACCTTCTTACACATAGAACGCTCTCCTACCATGA
	12019	TGCGTTCCAGCGCTCCTCACGGTACACCTTCTTACACATAGAACGCTCTCCTACCATGA
	12021	TGCGTTCCAGCGCTCCTCACGGTACACCTTCTTACACATAGAACGCTCTCCTACCATGA
	12020	TGCGTTCCAGCGCTCCTCACGGTACACCTTCTTACACATAGAACGCTCTCCTACCATGA
30	12018	TGCGTTCCAGCGCTCCTCACGGTACACCTTCTTACACATAGAACGCTCTCCTACCATGA
	12022	TGCGTTCCAGCGCTCCTCACGGTACACCTTCTTACACATAGAACGCTCTCCTACCATGA

35	12023	CACTTTTGTGTATCCACAGCTTCGGTAATATGTTTTAGCCCCGGTACATTTTCGGCGCA
	12024	CACTTTTGTGTATCCACAGCTTCGGTAATATGTTTTAGCCCCGGTACATTTTCGGCGCA
	12019	CACTTTTGTGTATCCACAGCTTCGGTAATATGTTTTAGCCCCGGTACATTTTCGGCGCA
	12021	CACTTTTGTGTATCCACAGCTTCGGTAATATGTTTTAGCCCCGGTACATTTTCGGCGCA
	12020	CACTTTTGTGTATCCACAGCTTCGGTAATATGTTTTAGCCCCGGTACATTTTCGGCGCA
	12018	CACTTTTGTGTATCCACAGCTTCGGTAATATGTTTTAGCCCCGGTACATTTTCGGCGCA
40	12022	CACTTTTGTGTATCCACAGCTTCGGTAATATGTTTTAGCCCCGGTACATTTTCGGCGCA

45	12023	GGGTCACTCGACTAGTGAGCTATTACGCACTCTTTGAATGAATAGCTGCTTCTAAGCTAA
	12024	GGGTCACTCGACTAGTGAGCTATTACGCACTCTTTGAATGAATAGCTGCTTCTAAGCTAA
	12019	GGGTCACTCGACTAGTGAGCTATTACGCACTCTTTGAATGAATAGCTGCTTCTAAGCTAA
	12021	GGGTCACTCGACTAGTGAGCTATTACGCACTCTTTGAATGAATAGCTGCTTCTAAGCTAA
	12020	GGGTCACTCGACTAGTGAGCTATTACGCACTCTTTGAATGAATAGCTGCTTCTAAGCTAA
	12018	GGGTCACTCGACTAGTGAGCTATTACGCACTCTTTGAATGAATAGCTGCTTCTAAGCTAA
	12022	GGGTCACTCGACTAGTGAGCTATTACGCACTCTTTGAATGAATAGCTGCTTCTAAGCTAA

50	12023	CATCCTAGTTGTCTGTGCAACCCACATCCTTTTCCACTTAACATATATTTTGGGACCTT
	12024	CATCCTAGTTGTCTGTGCAACCCACATCCTTTTCCACTTAACATATATTTTGGGACCTT
	12019	CATCCTAGTTGTCTGTGCAACCCACATCCTTTTCCACTTAACATATATTTTGGGACCTT
	12021	CATCCTAGTTGTCTGTGCAACCCACATCCTTTTCCACTTAACATATATTTTGGGACCTT
55	12020	CATCCTAGTTGTCTGTGCAACCCACATCCTTTTCCACTTAACATATATTTTGGGACCTT
	12018	CATCCTAGTTGTCTGTGCAACCCACATCCTTTTCCACTTAACATATATTTTGGGACCTT
	12022	CATCCTAGTTGTCTGTGCAACCCACATCCTTTTCCACTTAACATATATTTTGGGACCTT

60	12023	AGCTGGTGGTCTGGGCTGTTTCCCTTTCGACTACGGATCTTAGCACTCGCAGTCTGACTG
	12024	AGCTGGTGGTCTGGGCTGTTTCCCTTTCGACTACGGATCTTAGCACTCGCAGTCTGACTG
	12019	AGCTGGTGGTCTGGGCTGTTTCCCTTTCGACTACGGATCTTAGCACTCGCAGTCTGACTG
	12021	AGCTGGTGGTCTGGGCTGTTTCCCTTTCGACTACGGATCTTAGCACTCGCAGTCTGACTG
	12020	AGCTGGTGGTCTGGGCTGTTTCCCTTTCGACTACGGATCTTAGCACTCGCAGTCTGACTG
65	12018	AGCTGGTGGTCTGGGCTGTTTCCCTTTCGACTACGGATCTTAGCACTCGCAGTCTGACTG
	12022	AGCTGGTGGTCTGGGCTGTTTCCCTTTCGACTACGGATCTTAGCACTCGCAGTCTGACTG

-2962-

5	12023	CCGATTATATCTCGTTGGCATTTCGGAGTTTATCTGAGATTGGTAATCCGGGATGGACCCC
	12024	CCGATTATATCTCGTTGGCATTTCGGAGTTTATCTGAGATTGGTAATCCGGGATGGACCCC
	12019	CCGATTATATCTCGTTGGCATTTCGGAGTTTATCTGAGATTGGTAATCCGGGATGGACCCC
	12021	CCGATTATATCTCGTTGGCATTTCGGAGTTTATCTGAGATTGGTAATCCGGGATGGACCCC
	12020	CCGATTATATCTCGTTGGCATTTCGGAGTTTATCTGAGATTGGTAATCCGGGATGGACCCC
	12018	CCGATTATATCTCGTTGGCATTTCGGAGTTTATCTGAGATTGGTAATCCGGGATGGACCCC
10	12022	CCGATTATATCTCGTTGGCATTTCGGAGTTTATCTGAGATTGGTAATCCGGGATGGACCCC *****
	12023	TCACCCAAACAGTGCTCTACCTCCAAGAGACTTAACATCGACGCTAGCCCTAAAGCTATT
	12024	TCACCCAAACAGTGCTCTACCTCCAAGAGACTTAACATCGACGCTAGCCCTAAAGCTATT
	12019	TCACCCAAACAGTGCTCTACCTCCAAGAGACTTAACATCGACGCTAGCCCTAAAGCTATT
	12021	TCACCCAAACAGTGCTCTACCTCCAAGAGACTTAACATCGACGCTAGCCCTAAAGCTATT
	12020	TCACCCAAACAGTGCTCTACCTCCAAGAGACTTAACATCGACGCTAGCCCTAAAGCTATT
15	12018	TCACCCAAACAGTGCTCTACCTCCAAGAGACTTAACATCGACGCTAGCCCTAAAGCTATT
	12022	TCACCCAAACAGTGCTCTACCTCCAAGAGACTTAACATCGACGCTAGCCCTAAAGCTATT *****
20	12023	TCGGAGAGAACCAGCTATCTCCAAGTTCGTTTGGAAATTTCTCCGCTACCCACAAGTCATC
	12024	TCGGAGAGAACCAGCTATCTCCAAGTTCGTTTGGAAATTTCTCCGCTACCCACAAGTCATC
	12019	TCGGAGAGAACCAGCTATCTCCAAGTTCGTTTGGAAATTTCTCCGCTACCCACAAGTCATC
	12021	TCGGAGAGAACCAGCTATCTCCAAGTTCGTTTGGAAATTTCTCCGCTACCCACAAGTCATC
	12020	TCGGAGAGAACCAGCTATCTCCAAGTTCGTTTGGAAATTTCTCCGCTACCCACAAGTCATC
	12018	TCGGAGAGAACCAGCTATCTCCAAGTTCGTTTGGAAATTTCTCCGCTACCCACAAGTCATC
25	12022	TCGGAGAGAACCAGCTATCTCCAAGTTCGTTTGGAAATTTCTCCGCTACCCACAAGTCATC *****
30	12023	CAAGCACTTTTCAACGTGCCCTGGTTCGGTCCTCCAGTGAGTTTACCTCACCTTCAACC
	12024	CAAGCACTTTTCAACGTGCCCTGGTTCGGTCCTCCAGTGAGTTTACCTCACCTTCAACC
	12019	CAAGCACTTTTCAACGTGCCCTGGTTCGGTCCTCCAGTGAGTTTACCTCACCTTCAACC
	12021	CAAGCACTTTTCAACGTGCCCTGGTTCGGTCCTCCAGTGAGTTTACCTCACCTTCAACC
	12020	CAAGCACTTTTCAACGTGCCCTGGTTCGGTCCTCCAGTGAGTTTACCTCACCTTCAACC
	12018	CAAGCACTTTTCAACGTGCCCTGGTTCGGTCCTCCAGTGAGTTTACCTCACCTTCAACC
35	12022	CAAGCACTTTTCAACGTGCCCTGGTTCGGTCCTCCAGTGAGTTTACCTCACCTTCAACC *****
40	12023	TGCTCATGGGTAGGTACATGGTTTCGGGTCTACAACATGATACTATGACGCCCTATTAA
	12024	TGCTCATGGGTAGGTACATGGTTTCGGGTCTACAACATGATACTATGACGCCCTATTAA
	12019	TGCTCATGGGTAGGTACATGGTTTCGGGTCTACAACATGATACTATGACGCCCTATTAA
	12021	TGCTCATGGGTAGGTACATGGTTTCGGGTCTACAACATGATACTATGACGCCCTATTAA
	12020	TGCTCATGGGTAGGTACATGGTTTCGGGTCTACAACATGATACTATGACGCCCTATTAA
	12018	TGCTCATGGGTAGGTACATGGTTTCGGGTCTACAACATGATACTATGACGCCCTATTAA
45	12022	TGCTCATGGGTAGGTACATGGTTTCGGGTCTACAACATGATACTATGACGCCCTATTAA *****
50	12023	GACTCGGTTTCCCTACGGCTCCGTCTCTTCAACTTAACCTCGCATCATATCGTAACTCGC
	12024	GACTCGGTTTCCCTACGGCTCCGTCTCTTCAACTTAACCTCGCATCATATCGTAACTCGC
	12019	GACTCGGTTTCCCTACGGCTCCGTCTCTTCAACTTAACCTCGCATCATATCGTAACTCGC
	12021	GACTCGGTTTCCCTACGGCTCCGTCTCTTCAACTTAACCTCGCATCATATCGTAACTCGC
	12020	GACTCGGTTTCCCTACGGCTCCGTCTCTTCAACTTAACCTCGCATCATATCGTAACTCGC
	12018	GACTCGGTTTCCCTACGGCTCCGTCTCTTCAACTTAACCTCGCATCATATCGTAACTCGC
55	12022	GACTCGGTTTCCCTACGGCTCCGTCTCTTCAACTTAACCTCGCATCATATCGTAACTCGC *****
60	12023	CGGTTTCACTTCTACAAAAGGCACGCTCTCACCATTAAACGGGCTCGAACTTGTGTAGGCA
	12024	CGGTTTCACTTCTACAAAAGGCACGCTCTCACCATTAAACGGGCTCGAACTTGTGTAGGCA
	12019	CGGTTTCACTTCTACAAAAGGCACGCTCTCACCATTAAACGGGCTCGAACTTGTGTAGGCA
	12021	CGGTTTCACTTCTACAAAAGGCACGCTCTCACCATTAAACGGGCTCGAACTTGTGTAGGCA
	12020	CGGTTTCACTTCTACAAAAGGCACGCTCTCACCATTAAACGGGCTCGAACTTGTGTAGGCA
	12018	CGGTTTCACTTCTACAAAAGGCACGCTCTCACCATTAAACGGGCTCGAACTTGTGTAGGCA
65	12022	CGGTTTCACTTCTACAAAAGGCACGCTCTCACCATTAAACGGGCTCGAACTTGTGTAGGCA *****
65	12023	CACGGTTTCAGGTTCTATTTCACTCCCTCCCGGGTGCTTTTACCTTTCCCTCACGGT
	12024	CACGGTTTCAGGTTCTATTTCACTCCCTCCCGGGTGCTTTTACCTTTCCCTCACGGT
	12019	CACGGTTTCAGGTTCTATTTCACTCCCTCCCGGGTGCTTTTACCTTTCCCTCACGGT
	12021	CACGGTTTCAGGTTCTATTTCACTCCCTCCCGGGTGCTTTTACCTTTCCCTCACGGT
	12020	CACGGTTTCAGGTTCTATTTCACTCCCTCCCGGGTGCTTTTACCTTTCCCTCACGGT
	12020	CACGGTTTCAGGTTCTATTTCACTCCCTCCCGGGTGCTTTTACCTTTCCCTCACGGT

-2963-

	12018	CACGGTTTCAGGTTCTATTTCACTCCCCTCCCGGGTGCTTTTCACCTTTCCCTCACGGT
	12022	CACGGTTTCAGGTTCTATTTCACTCCCCTCCCGGGTGCTTTTCACCTTTCCCTCACGGT

5	12023	ACTGGTTCACATATCGGTCACTAGAGAGTATTTAGGGTTGGGAGATGGTCCTCCAGATTTC
	12024	ACTGGTTCACATATCGGTCACTAGAGAGTATTTAGGGTTGGGAGATGGTCCTCCAGATTTC
	12019	ACTGGTTCACATATCGGTCACTAGAGAGTATTTAGGGTTGGGAGATGGTCCTCCAGATTTC
	12021	ACTGGTTCACATATCGGTCACTAGAGAGTATTTAGGGTTGGGAGATGGTCCTCCAGATTTC
	12020	ACTGGTTCACATATCGGTCACTAGAGAGTATTTAGGGTTGGGAGATGGTCCTCCAGATTTC
10	12018	ACTGGTTCACATATCGGTCACTAGAGAGTATTTAGGGTTGGGAGATGGTCCTCCAGATTTC
	12022	ACTGGTTCACATATCGGTCACTAGAGAGTATTTAGGGTTGGGAGATGGTCCTCCAGATTTC

	12023	CGACGAGATTTTCGCGTGTCTCGCGTACTCAGGATACTGCTAAGGTTAATCTATCATTTT
15	12024	CGACGAGATTTTCGCGTGTCTCGCGTACTCAGGATACTGCTAAGGTTAATCTATCATTTT
	12019	CGACGAGATTTTCGCGTGTCTCGCGTACTCAGGATACTGCTAAGGTTAATCTATCATTTT
	12021	CGACGAGATTTTCGCGTGTCTCGCGTACTCAGGATACTGCTAAGGTTAATCTATCATTTT
	12020	CGACGAGATTTTCGCGTGTCTCGCGTACTCAGGATACTGCTAAGGTTAATCTATCATTTT
	12018	CGACGAGATTTTCGCGTGTCTCGCGTACTCAGGATACTGCTAAGGTTAATCTATCATTTT
20	12022	CGACGAGATTTTCGCGTGTCTCGCGTACTCAGGATACTGCTAAGGTTAATCTATCATTTT

	12023	AAATACGAGGCTGTTACTCTCTTTGGCTTACCTTCCCAGGTAATTCTTCTATAATGATTA
	12024	AAATACGAGGCTGTTACTCTCTTTGGCTTACCTTCCCAGGTAATTCTTCTATAATGATTA
25	12019	AAATACGAGGCTGTTACTCTCTTTGGCTTACCTTCCCAGGTAATTCTTCTATAATGATTA
	12021	AAATACGAGGCTGTTACTCTCTTTGGCTTACCTTCCCAGGTAATTCTTCTATAATGATTA
	12020	AAATACGAGGCTGTTACTCTCTTTGGCTTACCTTCCCAGGTAATTCTTCTATAATGATTA
	12018	AAATACGAGGCTGTTACTCTCTTTGGCTTACCTTCCCAGGTAATTCTTCTATAATGATTA
30	12022	AAATACGAGGCTGTTACTCTCTTTGGCTTACCTTCCCAGGTAATTCTTCTATAATGATTA

	12023	ATCCTATATCGCAGTCTTACAACCCCGAAGTGTAACACTTCGGTTTGCCTCCTGCCGT
	12024	ATCCTATATCGCAGTCTTACAACCCCGAAGTGTAACACTTCGGTTTGCCTCCTGCCGT
35	12019	ATCCTATATCGCAGTCTTACAACCCCGAAGTGTAACACTTCGGTTTGCCTCCTGCCGT
	12021	ATCCTATATCGCAGTCTTACAACCCCGAAGTGTAACACTTCGGTTTGCCTCCTGCCGT
	12020	ATCCTATATCGCAGTCTTACAACCCCGAAGTGTAACACTTCGGTTTGCCTCCTGCCGT
	12018	ATCCTATATCGCAGTCTTACAACCCCGAAGTGTAACACTTCGGTTTGCCTCCTGCCGT
	12022	ATCCTATATCGCAGTCTTACAACCCCGAAGTGTAACACTTCGGTTTGCCTCCTGCCGT

40		
	12023	TTCGCTCGCGCTACTAAGGCAATCGCTTTTGCTTTCTCTTCTGCGCTACTTAGATGT
	12024	TTCGCTCGCGCTACTAAGGCAATCGCTTTTGCTTTCTCTTCTGCGCTACTTAGATGT
	12019	TTCGCTCGCGCTACTAAGGCAATCGCTTTTGCTTTCTCTTCTGCGCTACTTAGATGT
	12021	TTCGCTCGCGCTACTAAGGCAATCGCTTTTGCTTTCTCTTCTGCGCTACTTAGATGT
45	12020	TTCGCTCGCGCTACTAAGGCAATCGCTTTTGCTTTCTCTTCTGCGCTACTTAGATGT
	12018	TTCGCTCGCGCTACTAAGGCAATCGCTTTTGCTTTCTCTTCTGCGCTACTTAGATGT
	12022	TTCGCTCGCGCTACTAAGGCAATCGCTTTTGCTTTCTCTTCTGCGCTACTTAGATGT

50	12023	TTCAGTTCACATGCGTCTTCTCTCATATCCTTAAACAGATATGGATACTAGTCATTAAC
	12024	TTCAGTTCACATGCGTCTTCTCTCATATCCTTAAACAGATATGGATACTAGTCATTAAC
	12019	TTCAGTTCACATGCGTCTTCTCTCATATCCTTAAACAGATATGGATACTAGTCATTAAC
	12021	TTCAGTTCACATGCGTCTTCTCTCATATCCTTAAACAGATATGGATACTAGTCATTAAC
	12020	TTCAGTTCACATGCGTCTTCTCTCATATCCTTAAACAGATATGGATACTAGTCATTAAC
55	12018	TTCAGTTCACATGCGTCTTCTCTCATATCCTTAAACAGATATGGATACTAGTCATTAAC
	12022	TTCAGTTCACATGCGTCTTCTCTCATATCCTTAAACAGATATGGATACTAGTCATTAAC

60	12023	AGTGGGTTCCCCCATTCGGACATCTCTGGATCAGCGCTTACTTACAGCTCCCCAAGCAT
	12024	AGTGGGTTCCCCCATTCGGACATCTCTGGATCAGCGCTTACTTACAGCTCCCCAAGCAT
	12019	AGTGGGTTCCCCCATTCGGACATCTCTGGATCAGCGCTTACTTACAGCTCCCCAAGCAT
	12021	AGTGGGTTCCCCCATTCGGACATCTCTGGATCAGCGCTTACTTACAGCTCCCCAAGCAT
	12020	AGTGGGTTCCCCCATTCGGACATCTCTGGATCAGCGCTTACTTACAGCTCCCCAAGCAT
	12018	AGTGGGTTCCCCCATTCGGACATCTCTGGATCAGCGCTTACTTACAGCTCCCCAAGCAT
65	12022	AGTGGGTTCCCCCATTCGGACATCTCTGGATCAGCGCTTACTTACAGCTCCCCAAGCAT

	12023	TTCTGCTTAGTCACGTCCTTCTTCTCGGCTTCTAGTGCCCAAGGCATCCACCGTGCGCCCTT

-2964-

5	12024	TTCTGTCGTTAGTCACGTCCTTCTTCGGCTTCTAGTGCCAAGGCATCCACCGTGCGCCCTT
	12019	TTCTGTCGTTAGTCACGTCCTTCTTCGGCTTCTAGTGCCAAGGCATCCACCGTGCGCCCTT
	12021	TTCTGTCGTTAGTCACGTCCTTCTTCGGCTTCTAGTGCCAAGGCATCCACCGTGCGCCCTT
	12020	TTCTGTCGTTAGTCACGTCCTTCTTCGGCTTCTAGTGCCAAGGCATCCACCGTGCGCCCTT
	12018	TTCTGTCGTTAGTCACGTCCTTCTTCGGCTTCTAGTGCCAAGGCATCCACCGTGCGCCCTT
10	12022	TTCTGTCGTTAGTCACGTCCTTCTTCGGCTTCTAGTGCCAAGGCATCCACCGTGCGCCCTT *****
	12023	ATTAACTTAACCTTATTAACTAGTTTCTTTAAAACTAGAAAACTCATTAAATATTTCACA
	12024	ATTAACTTAACCTTATTAACTAGTTTCTTTAAAACTAGAAAACTCATTAAATATTTCACA
	12019	ATTAACTTAACCTTATTAACTAGTTTCTTTAAAACTAGAAAACTCATTAAATATTTCACA
	12021	ATTAACTTAACCTTATTAACTAGTTTCTTTAAAACTAGAAAACTCATTAAATATTTCACA
15	12020	ATTAACTTAACCTTATTAACTAGTTTCTTTAAAACTAGAAAACTCATTAAATATTTCACA
	12018	ATTAACTTAACCTTATTAACTAGTTTCTTTAAAACTAGAAAACTCATTAAATATTTCACA
	12022	ATTAACTTAACCTTATTAACTAGTTTCTTTAAAACTAGAAAACTCATTAAATATTTCACA *****
20	12023	GCGTTTTTCGGTTTATTTTCTGTACTTTCTACAATCTATTTCTAGATCGTGGAATTTGA
	12024	GCGTTTTTCGGTTTATTTTCTGTACTTTCTACAATCTATTTCTAGATCGTGGAATTTGA
	12019	GCGTTTTTCGGTTTATTTTCTGTACTTTCTACAATCTATTTCTAGATCGTGGAATTTGA
	12021	GCGTTTTTCGGTTTATTTTCTGTACTTTCTACAATCTATTTCTAGATCGTGGAATTTGA
	12020	GCGTTTTTCGGTTTATTTTCTGTACTTTCTACAATCTATTTCTAGATCGTGGAATTTGA
25	12018	GCGTTTTTCGGTTTATTTTCTGTACTTTCTACAATCTATTTCTAGATCGTGGAATTTGA
	12022	GCGTTTTTCGGTTTATTTTCTGTACTTTCTACAATCTATTTCTAGATCGTGGAATTTGA *****
30	12023	TATAGATATTCAATTTTCAATGAACAATTGAACCTTTTCGATTCAATGGAGCCTAGCGGG
	12024	TATAGATATTCAATTTTCAATGAACAATTGAACCTTTTCGATTCAATGGAGCCTAGCGGG
	12019	TATAGATATTCAATTTTCAATGAACAATTGAACCTTTTCGATTCAATGGAGCCTAGCGGG
	12021	TATAGATATTCAATTTTCAATGAACAATTGAACCTTTTCGATTCAATGGAGCCTAGCGGG
	12020	TATAGATATTCAATTTTCAATGAACAATTGAACCTTTTCGATTCAATGGAGCCTAGCGGG
35	12018	TATAGATATTCAATTTTCAATGAACAATTGAACCTTTTCGATTCAATGGAGCCTAGCGGG
	12022	TATAGATATTCAATTTTCAATGAACAATTGAACCTTTTCGATTCAATGGAGCCTAGCGGG *****
40	12023	ATCGAACCGCTGACCTCTGCGTGCAAAGCAGGCGCTCTCCAGCTGAGCTAAGGCCCA
	12024	ATCGAACCGCTGACCTCTGCGTGCAAAGCAGGCGCTCTCCAGCTGAGCTAAGGCCCA
	12019	ATCGAACCGCTGACCTCTGCGTGCAAAGCAGGCGCTCTCCAGCTGAGCTAAGGCCCA
	12021	ATCGAACCGCTGACCTCTGCGTGCAAAGCAGGCGCTCTCCAGCTGAGCTAAGGCCCA
	12020	ATCGAACCGCTGACCTCTGCGTGCAAAGCAGGCGCTCTCCAGCTGAGCTAAGGCCCA
45	12018	ATCGAACCGCTGACCTCTGCGTGCAAAGCAGGCGCTCTCCAGCTGAGCTAAGGCCCA
	12022	ATCGAACCGCTGACCTCTGCGTGCAAAGCAGGCGCTCTCCAGCTGAGCTAAGGCCCA *****
50	12023	CAAGACCTCTCAAACTAAACAAGACGCAAATGGCAGGTTTCCTTATCCTTAGAAAGGAG
	12024	CAAGACCTCTCAAACTAAACAAGACGCAAATGGCAGGTTTCCTTATCCTTAGAAAGGAG
	12019	CAAGACCTCTCAAACTAAACAAGACGCAAATGGCAGGTTTCCTTATCCTTAGAAAGGAG
	12021	CAAGACCTCTCAAACTAAACAAGACGCAAATGGCAGGTTTCCTTATCCTTAGAAAGGAG
	12020	CAAGACCTCTCAAACTAAACAAGACGCAAATGGCAGGTTTCCTTATCCTTAGAAAGGAG
55	12018	CAAGACCTCTCAAACTAAACAAGACGCAAATGGCAGGTTTCCTTATCCTTAGAAAGGAG
	12022	CAAGACCTCTCAAACTAAACAAGACGCAAATGGCAGGTTTCCTTATCCTTAGAAAGGAG *****
60	12023	GTGATCCAGCCGACCTTCCGATACGGCTACCTTGTACGACTTCACCCCAATCATCTAT
	12024	GTGATCCAGCCGACCTTCCGATACGGCTACCTTGTACGACTTCACCCCAATCATCTAT
	12019	GTGATCCAGCCGACCTTCCGATACGGCTACCTTGTACGACTTCACCCCAATCATCTAT
	12021	GTGATCCAGCCGACCTTCCGATACGGCTACCTTGTACGACTTCACCCCAATCATCTAT
	12020	GTGATCCAGCCGACCTTCCGATACGGCTACCTTGTACGACTTCACCCCAATCATCTAT
65	12018	GTGATCCAGCCGACCTTCCGATACGGCTACCTTGTACGACTTCACCCCAATCATCTAT
	12022	GTGATCCAGCCGACCTTCCGATACGGCTACCTTGTACGACTTCACCCCAATCATCTAT *****
	12023	CCCACCTTAGGCGGCTGGCTCCTAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTC
	12024	CCCACCTTAGGCGGCTGGCTCCTAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTC
	12019	CCCACCTTAGGCGGCTGGCTCCTAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTC
	12021	CCCACCTTAGGCGGCTGGCTCCTAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTC
	12020	CCCACCTTAGGCGGCTGGCTCCTAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTC
	12018	CCCACCTTAGGCGGCTGGCTCCTAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTC
	12018	CCCACCTTAGGCGGCTGGCTCCTAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTC

-2965-

	12022	CCCACCTTAGGCGGCTGGCTCCTAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTC *****
5	12023	TCCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTACCGCGGCGTGCTGAT
	12024	TCCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTACCGCGGCGTGCTGAT
	12019	TCCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTACCGCGGCGTGCTGAT
	12021	TCCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTACCGCGGCGTGCTGAT
	12020	TCCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTACCGCGGCGTGCTGAT
10	12018	TCCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTACCGCGGCGTGCTGAT
	12022	TCCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTACCGCGGCGTGCTGAT *****
	12023	CCGCGATTACTAGCGATTCCGACTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACCTGA
15	12024	CCGCGATTACTAGCGATTCCGACTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACCTGA
	12019	CCGCGATTACTAGCGATTCCGACTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACCTGA
	12021	CCGCGATTACTAGCGATTCCGACTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACCTGA
	12020	CCGCGATTACTAGCGATTCCGACTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACCTGA
	12018	CCGCGATTACTAGCGATTCCGACTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACCTGA
20	12022	CCGCGATTACTAGCGATTCCGACTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACCTGA *****
	12023	GATTGGCTTTAAGAGATTAGCTTGCCGTCACCGGCTTGCGACTCGTTGTACCAACCATTG
25	12024	GATTGGCTTTAAGAGATTAGCTTGCCGTCACCGGCTTGCGACTCGTTGTACCAACCATTG
	12019	GATTGGCTTTAAGAGATTAGCTTGCCGTCACCGGCTTGCGACTCGTTGTACCAACCATTG
	12021	GATTGGCTTTAAGAGATTAGCTTGCCGTCACCGGCTTGCGACTCGTTGTACCAACCATTG
	12020	GATTGGCTTTAAGAGATTAGCTTGCCGTCACCGGCTTGCGACTCGTTGTACCAACCATTG
	12018	GATTGGCTTTAAGAGATTAGCTTGCCGTCACCGGCTTGCGACTCGTTGTACCAACCATTG
30	12022	GATTGGCTTTAAGAGATTAGCTTGCCGTCACCGGCTTGCGACTCGTTGTACCAACCATTG *****
	12023	TAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCC
35	12024	TAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCC
	12019	TAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCC
	12021	TAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCC
	12020	TAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCC
	12018	TAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCC
	12022	TAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCC *****
40	12023	TCCGGTTTATTACCGGCAGTCTCGCTAGAGTGCCCAACTTAATGATGGCAACTAACAAATA
	12024	TCCGGTTTATTACCGGCAGTCTCGCTAGAGTGCCCAACTTAATGATGGCAACTAACAAATA
	12019	TCCGGTTTATTACCGGCAGTCTCGCTAGAGTGCCCAACTTAATGATGGCAACTAACAAATA
	12021	TCCGGTTTATTACCGGCAGTCTCGCTAGAGTGCCCAACTTAATGATGGCAACTAACAAATA
45	12020	TCCGGTTTATTACCGGCAGTCTCGCTAGAGTGCCCAACTTAATGATGGCAACTAACAAATA
	12018	TCCGGTTTATTACCGGCAGTCTCGCTAGAGTGCCCAACTTAATGATGGCAACTAACAAATA
	12022	TCCGGTTTATTACCGGCAGTCTCGCTAGAGTGCCCAACTTAATGATGGCAACTAACAAATA *****
	12023	GGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
50	12024	GGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
	12019	GGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
	12021	GGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
	12020	GGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
55	12018	GGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
	12022	GGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA *****
	12023	TGCACCACCTGTCACTTCTGCTCCGAAGAGAAAGCCTATCTCTAGGCCGGTCAGAAGGAT
60	12024	TGCACCACCTGTCACTTCTGCTCCGAAGAGAAAGCCTATCTCTAGGCCGGTCAGAAGGAT
	12019	TGCACCACCTGTCACTTCTGCTCCGAAGAGAAAGCCTATCTCTAGGCCGGTCAGAAGGAT
	12021	TGCACCACCTGTCACTTCTGCTCCGAAGAGAAAGCCTATCTCTAGGCCGGTCAGAAGGAT
	12020	TGCACCACCTGTCACTTCTGCTCCGAAGAGAAAGCCTATCTCTAGGCCGGTCAGAAGGAT
	12018	TGCACCACCTGTCACTTCTGCTCCGAAGAGAAAGCCTATCTCTAGGCCGGTCAGAAGGAT
65	12022	TGCACCACCTGTCACTTCTGCTCCGAAGAGAAAGCCTATCTCTAGGCCGGTCAGAAGGAT *****
	12023	GTCAAGACCTGGTAAGGTTCTTCGCGTGTGCTTGAATTAACCAACATGCTCCACCGCTTG
	12024	GTCAAGACCTGGTAAGGTTCTTCGCGTGTGCTTGAATTAACCAACATGCTCCACCGCTTG

-2966-

5	12019	GTCAAGACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTG
	12021	GTCAAGACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTG
	12020	GTCAAGACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTG
	12018	GTCAAGACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTG
	12022	GTCAAGACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTG *****
10	12023	TGCGGGCCCCGTCAATTCTTTGAGTTTCAACCTTGCGGTGCTACTCCCCAGGCGGAGT
	12024	TGCGGGCCCCGTCAATTCTTTGAGTTTCAACCTTGCGGTGCTACTCCCCAGGCGGAGT
	12019	TGCGGGCCCCGTCAATTCTTTGAGTTTCAACCTTGCGGTGCTACTCCCCAGGCGGAGT
	12021	TGCGGGCCCCGTCAATTCTTTGAGTTTCAACCTTGCGGTGCTACTCCCCAGGCGGAGT
	12020	TGCGGGCCCCGTCAATTCTTTGAGTTTCAACCTTGCGGTGCTACTCCCCAGGCGGAGT
15	12018	TGCGGGCCCCGTCAATTCTTTGAGTTTCAACCTTGCGGTGCTACTCCCCAGGCGGAGT
	12022	TGCGGGCCCCGTCAATTCTTTGAGTTTCAACCTTGCGGTGCTACTCCCCAGGCGGAGT *****
20	12023	GCTTAATGCGTTAGCTGCGGCACCTAAGCCCCGAAAGGGCCTAACACCTAGCACTCATCG
	12024	GCTTAATGCGTTAGCTGCGGCACCTAAGCCCCGAAAGGGCCTAACACCTAGCACTCATCG
	12019	GCTTAATGCGTTAGCTGCGGCACCTAAGCCCCGAAAGGGCCTAACACCTAGCACTCATCG
	12021	GCTTAATGCGTTAGCTGCGGCACCTAAGCCCCGAAAGGGCCTAACACCTAGCACTCATCG
	12020	GCTTAATGCGTTAGCTGCGGCACCTAAGCCCCGAAAGGGCCTAACACCTAGCACTCATCG
25	12018	GCTTAATGCGTTAGCTGCGGCACCTAAGCCCCGAAAGGGCCTAACACCTAGCACTCATCG
	12022	GCTTAATGCGTTAGCTGCGGCACCTAAGCCCCGAAAGGGCCTAACACCTAGCACTCATCG *****
30	12023	TTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCA
	12024	TTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCA
	12019	TTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCA
	12021	TTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCA
	12020	TTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCA
35	12018	TTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCA
	12022	TTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCA *****
40	12023	GCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCTCCATATATCTACGCA
	12024	GCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCTCCATATATCTACGCA
	12019	GCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCTCCATATATCTACGCA
	12021	GCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCTCCATATATCTACGCA
	12020	GCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCTCCATATATCTACGCA
45	12018	GCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCTCCATATATCTACGCA
	12022	GCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCTCCATATATCTACGCA *****
50	12023	TTTACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTCCTCCAGTTTCCA
	12024	TTTACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTCCTCCAGTTTCCA
	12019	TTTACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTCCTCCAGTTTCCA
	12021	TTTACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTCCTCCAGTTTCCA
	12020	TTTACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTCCTCCAGTTTCCA
55	12018	TTTACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTCCTCCAGTTTCCA
	12022	TTTACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTCCTCCAGTTTCCA *****
60	12023	AAGCGTACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTAAAGAACCGCCTGCGCTCG
	12024	AAGCGTACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTAAAGAACCGCCTGCGCTCG
	12019	AAGCGTACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTAAAGAACCGCCTGCGCTCG
	12021	AAGCGTACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTAAAGAACCGCCTGCGCTCG
	12020	AAGCGTACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTAAAGAACCGCCTGCGCTCG
65	12018	AAGCGTACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTAAAGAACCGCCTGCGCTCG
	12022	AAGCGTACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTAAAGAACCGCCTGCGCTCG *****
	12023	CTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCTGCTGGCA
	12024	CTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCTGCTGGCA
	12019	CTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCTGCTGGCA
	12021	CTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCTGCTGGCA
	12020	CTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCTGCTGGCA
	12018	CTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCTGCTGGCA
	12022	CTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCTGCTGGCA

-2967-

5	12023	CGTAGTTAGCCGTCCTTTCTGGTTAGTTACCGTCACCTGGTAGATTTTCCACTCCTACC
	12024	CGTAGTTAGCCGTCCTTTCTGGTTAGTTACCGTCACCTGGTAGATTTTCCACTCCTACC
	12019	CGTAGTTAGCCGTCCTTTCTGGTTAGTTACCGTCACCTGGTAGATTTTCCACTCCTACC
	12021	CGTAGTTAGCCGTCCTTTCTGGTTAGTTACCGTCACCTGGTAGATTTTCCACTCCTACC
	12020	CGTAGTTAGCCGTCCTTTCTGGTTAGTTACCGTCACCTGGTAGATTTTCCACTCCTACC
10	12018	CGTAGTTAGCCGTCCTTTCTGGTTAGTTACCGTCACCTGGTAGATTTTCCACTCCTACC
	12022	CGTAGTTAGCCGTCCTTTCTGGTTAGTTACCGTCACCTGGTAGATTTTCCACTCCTACC

15	12023	AACGTTCTTCTCTAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGT
	12024	AACGTTCTTCTCTAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGT
	12019	AACGTTCTTCTCTAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGT
	12021	AACGTTCTTCTCTAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGT
	12020	AACGTTCTTCTCTAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGT
20	12018	AACGTTCTTCTCTAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGT
	12022	AACGTTCTTCTCTAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGT

25	12023	TGCTCGGTGAGACTTCCGTCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTC
	12024	TGCTCGGTGAGACTTCCGTCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTC
	12019	TGCTCGGTGAGACTTCCGTCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTC
	12021	TGCTCGGTGAGACTTCCGTCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTC
	12020	TGCTCGGTGAGACTTCCGTCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTC
30	12018	TGCTCGGTGAGACTTCCGTCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTC
	12022	TGCTCGGTGAGACTTCCGTCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTC

35	12023	TGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCTCTCAGGTGGGTATGTATCGTCG
	12024	TGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCTCTCAGGTGGGTATGTATCGTCG
	12019	TGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCTCTCAGGTGGGTATGTATCGTCG
	12021	TGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCTCTCAGGTGGGTATGTATCGTCG
	12020	TGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCTCTCAGGTGGGTATGTATCGTCG
40	12018	TGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCTCTCAGGTGGGTATGTATCGTCG
	12022	TGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCTCTCAGGTGGGTATGTATCGTCG

45	12023	CCTTGGTGAGCCTTTACCTCACCAACTAGCTAATAACAACGAGGTCCATCTCACAGTGAA
	12024	CCTTGGTGAGCCTTTACCTCACCAACTAGCTAATAACAACGAGGTCCATCTCACAGTGAA
	12019	CCTTGGTGAGCCTTTACCTCACCAACTAGCTAATAACAACGAGGTCCATCTCACAGTGAA
	12021	CCTTGGTGAGCCTTTACCTCACCAACTAGCTAATAACAACGAGGTCCATCTCACAGTGAA
	12020	CCTTGGTGAGCCTTTACCTCACCAACTAGCTAATAACAACGAGGTCCATCTCACAGTGAA
50	12018	CCTTGGTGAGCCTTTACCTCACCAACTAGCTAATAACAACGAGGTCCATCTCACAGTGAA
	12022	CCTTGGTGAGCCTTTACCTCACCAACTAGCTAATAACAACGAGGTCCATCTCACAGTGAA

55	12023	GCAATGCTCCTTTTAAATAACTAACATGTGTTAATTACTCTTATGCGGTATTAGCTATC
	12024	GCAATGCTCCTTTTAAATAACTAACATGTGTTAATTACTCTTATGCGGTATTAGCTATC
	12019	GCAATGCTCCTTTTAAATAACTAACATGTGTTAATTACTCTTATGCGGTATTAGCTATC
	12021	GCAATGCTCCTTTTAAATAACTAACATGTGTTAATTACTCTTATGCGGTATTAGCTATC
	12020	GCAATGCTCCTTTTAAATAACTAACATGTGTTAATTACTCTTATGCGGTATTAGCTATC
60	12018	GCAATGCTCCTTTTAAATAACTAACATGTGTTAATTACTCTTATGCGGTATTAGCTATC
	12022	GCAATGCTCCTTTTAAATAACTAACATGTGTTAATTACTCTTATGCGGTATTAGCTATC

65	12023	GTTTCCAATAGTTATCCCCCGCTATGAGGCAGGTTACCTACGCGTTACTCACCCGTTTCGC
	12024	GTTTCCAATAGTTATCCCCCGCTATGAGGCAGGTTACCTACGCGTTACTCACCCGTTTCGC
	12019	GTTTCCAATAGTTATCCCCCGCTATGAGGCAGGTTACCTACGCGTTACTCACCCGTTTCGC
	12021	GTTTCCAATAGTTATCCCCCGCTATGAGGCAGGTTACCTACGCGTTACTCACCCGTTTCGC
	12020	GTTTCCAATAGTTATCCCCCGCTATGAGGCAGGTTACCTACGCGTTACTCACCCGTTTCGC
	12018	GTTTCCAATAGTTATCCCCCGCTATGAGGCAGGTTACCTACGCGTTACTCACCCGTTTCGC
	12022	GTTTCCAATAGTTATCCCCCGCTATGAGGCAGGTTACCTACGCGTTACTCACCCGTTTCGC

	12023	AACTCATCAGTCTAGTGTAACACCAAACCTCAGCGTTTCTACTTGCATGTATTAGGCACG
	12024	AACTCATCAGTCTAGTGTAACACCAAACCTCAGCGTTTCTACTTGCATGTATTAGGCACG
	12019	AACTCATCAGTCTAGTGTAACACCAAACCTCAGCGTTTCTACTTGCATGTATTAGGCACG

-2968-

12021 AACTCATCAGTCTAGTGTAAACACCAAACCTCAGCGTTCTACTTGCATGTATTAGGCACG
12020 AACTCATCAGTCTAGTGTAAACACCAAACCTCAGCGTTCTACTTGCATGTATTAGGCACG
12018 AACTCATCAGTCTAGTGTAAACACCAAACCTCAGCGTTCTACTTGCATGTATTAGGCACG
5 12022 AACTCATCAGTCTAGTGTAAACACCAAACCTCAGCGTTCTACTTGCATGTATTAGGCACG

12023 CCGCCAGCGTTCGTCTGAGCCAGGATCAAACCTCTCATTAAGTTTGGAGCTTTGCTCTT
12024 CCGCCAGCGTTCGTCTGAGCCAGGATCAAACCTCTCATTAAGTTTGGAGCTTTGCTCTT
10 12019 CCGCCAGCGTTCGTCTGAGCCAGGATCAAACCTCTCATTAAGTTTGGAGCTTTGCTCTT
12021 CCGCCAGCGTTCGTCTGAGCCAGGATCAAACCTCTCATTAAGTTTGGAGCTTTGCTCTT
12020 CCGCCAGCGTTCGTCTGAGCCAGGATCAAACCTCTCATTAAGTTTGGAGCTTTGCTCTT
12018 CCGCCAGCGTTCGTCTGAGCCAGGATCAAACCTCTCATTAAGTTTGGAGCTTTGCTCTT
12022 CCGCCAGCGTTCGTCTGAGCCAGGATCAAACCTCTCATTAAGTTTGGAGCTTTGCTCTT

15 12023 TTCTGTCTCGCTGACAGATTTATTGTTTTTTT-GTCATTGACGGATTTACAATGTAAATCC
12024 TTCTGTCTCGCTGACAGATTTATTGTTTTTTTGTCTGACGGATTTACAATGTAAATCC
12019 TTCTGTCTCGCTGACAGATTTATTGTTTTTTT-GTCATTGACGGATTTACAATGTAAATCC
20 12021 TTCTGTCTCGCTGACAGATTTATTGTTTTTTTGTCTGACGGATTTACAATGTAAATCC
12020 TTCTGTCTCGCTGACAGATTTATTGTTTTTTTGTCTGACGGATTTACAATGTAAATCC
12018 TTCTGTCTCGCTGACAGATTTATTGTTTTTTT-GTCATTGACGGATTTACAATGTAAATCC
12022 TTCTGTCTCGCTGACAGATTTATTGTTTTTTT-GTCATTGACGGATTTACAATGTAAATCC

25 12023 ACCCTGCACATTGTTTCGTCTTGTTCAGTTTTCAAAGGTCTAATGATATATCATAAAAT
12024 ACCCTGCACATTGTTTCGTCTTGTTCAGTTTTCAAAGGTCTTGCCTCTCTTGAGACAAC
12019 ACCCTGCACATTGTTTCGTCTTGTTCAGTTTTCAAAGGTCTTGCCTCTCTTGAGACAAC
30 12021 ACCCTGCACATTGTTTCGTCTTGTTCAGTTTTCAAAGGTCTTGCCTCTCTTGAGACAAC
12020 ACCCTGCACATTGTTTCGTCTTGTTCAGTTTTCAAAGGTCTTGCCTCTCTTGAGACAAC
12018 ACCCTGCACATTGTTTCGTCTTGTTCAGTTTTCAAAGGTCTTGCCTCTCTTGAGACAAC
12022 ACCCTGCACATTGTTTCGTCTTGTTCAGTTTTCAAAGGTCTTGCCTCTCTTGAGACAAC

35 12023 ATATCCATCGGAAGACAGGATTCGAACCTG-CGACACCTTGGTCCCACCAAGTACTC
12024 TTCTATATTCTAGCAAACCTTATTCTGCTTTGTCAACTACTTTT-TTTAAGTTGTTAACTA
12019 TTCTATATTCTAGCAAACCTTATTCTGCTTTGTCAACTACTTT-TTTAAGTTGTTAACTA
12021 TTCTATATTCTAGCAAACCTTATTCTGCTTTGTCAACTACTTT-TTTAAGTTGTTAACTA
12020 TTCTATATTCTAGCAAACCTTATTCTGCTTTGTCAACTACTTT-TTTAAGTTGTTAACTA
40 12018 TTCTATATTCTAGCAAACCTTATTCTGCTTTGTCAACTACTTT-TTTAAGTTGTTAACTA
12022 TTCTATATTCTAGCAAACCTTATTCTGCTTTGTCAACTACTTT-TTTAAGTTGTTAACTA
* * * * *
12023 TACCAAGCTG--A-GCTACT-TCCCGAAAA--TATGCACC--CTAGAGGAGTCTGAAC
45 12024 CGCGTTACTAGAA-GCTGCTCTCTCGAGACAACCTTATTATTACTAAATATTTCTACT
12019 CGCGCTAATAGAA-CTGCTCTCTCGAGACAACCTTATTATTACTAAATATTTCTACT
12021 CGCGCTAATAGAA-CTGCTCTCTCGAGACAACCTTATTAGTTTACTACATCATCTCTTA
12020 CGCGCTAATAGAA-CTGCTCTCTCGAGACAACCTTATTAGTTTACTACATCATCTCTTA
12018 AATGATAATACAATATTAGGTTCTGCTTAAGAACTCATTAGTATACTATAATTTTTATT
50 12022 CGCGCTAATAGAA-CTGCTCTCTCGAGACAACCTTATTAGTTTACTACATCATCTCTTA
* * * * *
12023 CTCTAACCGCTGATTCTGTA-GTCAGGTACTCTATCC-----AGTTGA---GCTAAG
12024 TCCTGTCAATACTATTTTTGCATTTTCTTTTATTTTTAAA-AAGTTAATAATTATTAT
55 12019 TCCTGTCAATACTATTTTTGTGA---TTTATAAAATTTAGTAT-AGACATAACTATTCTC
12021 CTTTGTCAACTCTTTTTTCTACT-TTTTCTACATTTTCTGA-AAATGTAGATCAGGCTC
12020 CTTTGTCAACTCTTTTTTCTACT-TTTTCTACATTTTCTGA-AAATGTAGATAGAGCGC
12018 TGTGTCAATAGGTTTAAAAA-----AATCTCAGAGAAAACCTGAGATTTT
12022 CTTTGTCAACTCTTTTTTCTACT-TTTTCTACATTTTCTGAAAAAGTTTCTGTTGGC
* * * * *
60 12023 GGTGCTAAAT-----ATTATATGCCGA-----GGACCGGAATC-----G---A
12024 AGTAACTAAC-----CTTCTATACTTGTGA-ATGGATAGCATTT-----T---T
12019 TATATTCAATTAAAGAGAAATATATAACCACTATTGAGAAATGTAGTC-----T---A
65 12021 AA-GCTTAAC---GATTCTTTTAAATCATT-----AATTTAAAA-----C---A
12020 AAGAAAAAAGAGTCTCACCTCTTTTATTCTTAGTAACTACTACA-----A---A
12018 TAAATT--ATGTTACAAAGTT--AATTTCTT-----TAGCTTCAATT-----AAA
12022 TAACACCAATAACATAGAGTTTAAATTCATAC--CTAAATTTATTTATTAGTAAAA
*

-2969-

5	12023	ACCGGTACGATGTTTACC-A--TCGCAGGATTTTAAGTCTGTGCGTCTGC--CAGTTCC
	12024	ACCGTTGTCATGTTTCAT--A--TTTCATCTTCTTAATTACAAATTTAAACTTCATCTTC
	12019	GCGATTAAATTCCTTGCTCA--TCGAA-AATATCCAAATAAATAATATGCATAAAACG
	12021	AATTTTCAGACATGTTGC-----CAAA-GTTTTGATATTATTACTATAAT--ATAGTTTG
	12020	TCTATTAGGATCGTTACCTT--CAGAATAACTTTCAACACCCTCTATAGT-TGCAATTGT
	12018	CCTAGTTCCGCATCTTCACG-CTGTAAAGGACATTTGTGCTATTATCTTCTGCATCT--
10	12022	AATAAAGATGGGCTAGCCATCTTTTAATAATTTGTTTTTTATATTCTTCAGCTTCTTG
		* *
15	12023	G-CCA-----CCCCGGCCTCTAACAAGCGAACGACGGGGTTCGAA-CCCCGACCCCT
	12024	A-TAAAAAATACCCCTTCAAATTTTATCTAAATTTGAAGGGTATTGAAATTTATAAAGTT
	12019	C-CTGCTTACGAAATATAACAAA-ATTGTTTGCAT--TTCGTAAACAAGCGTTACCTAT
	12021	T-AGAGGAGAATAATATGGGCCAA-GAACCTATCAT--CGAATATCAAAATAT----CAA
	12020	T-TTATGAACAGTTTTTTCGCTCACTGTTACTCATAGGATCCATATGGTAAGGTTCAATTAG
	12018	--GTATAGATAAAGAAATCATGACCTAAAAGTTCCATTTGCAACAATGCTTCTCTCAACAT
20	12022	GGGTGTAGATAAAACAAA-ATGACCAGGGTAATCTCGTGCATTTGACGTTCTTGTCCGT
25	12023	CAC--CTTGGCAAGGTGATGTTCTACCACTGAACTACGT-TCGCACTAAAGACACTATT-
	12024	CTT--TAAAAATATATGATGACTTATTTTATCTTCTCTTGCATTTTCTTTGATTT
	12019	TTA--ACAATATATGATGAGTGTTCCTCGCTGAGAATAATTCTCAGCGGTAGACCAGAGCT
	12021	TAA--AGTGTATGGGAAAATGTTGCGGTTGAAGACA---TTAACCTTAAAAATTACCC
	12020	TCT--CTAAACACGCGCTAGCTATTTTCTAGAAAAA---TCAATTAAAGTTTCTGTAC
	12018	CCA--TTGGTTTTAGATTAACTTCTTAGTACGTACAAT--T-CTTTGGCTTACTGCTTC
30	12022	CTTGCTCAATAGCTGGATTATACGGCTGGTGAACACGTT--GACGTTCACTCTCCGGATC
		*
35	12023	-----
	12024	CATCGTATGATAACGCTCTTGCTTTATCTTCA--TCATTTTCTGTCTCAGGCATTTTAC
	12019	AGACTAAGAAATCGATTGATTCATCATATAACACTCAACAAAATTGATAAAAATATAC
	12021	TGGT---GATTTTGTGTTTTCATCGGTA-----CGAGTGGATCAGGTAACAACAT
	12020	GATGCTCAACGTAGTCATGGACATTAATGGA---TACTGAAAACTCTAGAAAAGCGG
	12018	TTTCATCTGGCTCAGCC----TCAAATCTGTTGTGAAAAC---TTGACTTGCTGGAATC
40	12022	TGGTCTGGAATAGCTGATAATAGACTCTTCGTATAAGGGTGGATTGGATTGTTATAAAC
	12023	-----
	12024	CTGTCTCAAAAATCGATTTAATCTGAGCAGCATCAA----GAGTCTCATATTTTAAGAG
	12019	TAATTCA-ATAATTGCCATTGGGGCAGCATCGCCAC--GGCGTGGTTCGT-TTTAAGAA
	12021	TAATGCGTATGGTTAACCATATGTTAAACCAACAA--ATGGTACTCTATTATTAAAGG
	12020	TCATGAAGATAATT----TTGTGCTAACCACTGCAACGATTTTAATACTTT-TCCATGAT
45	12018	TTTTACGATATT-----TTTTCGCAATTTTA-----GTTTTATT-TTTACGA-
	12022	ATCATCAGATGTTT--CAACTTCTAACAGTTTCCCCCAATGCATAACACC-GATACGAT
50	12023	-----
	12024	GGCTTCTGCAATTAATTTATGAGTATCACGGTTTTCGTTGATAATATCAGCTGCC--TTA
	12019	TACGAGTGATCCTCCG--TTACGTTTACGATAACGAGGTGCGATGTCGTC-AAA--AAG
	12021	AAAAGATATTTCTACTA--TTAACCCCATTTGAATTAAGACGCAGAATTG--GAT--ATG
	12020	AACCAATGATGCGCCAGCTTCTGGCGTTTCTATTGGAGATTATTTGTC-GCT--TAC
	12018	-----ATTTGAC-----GCTCAATTTT-----ATCAACAATAAGTCAAT-----
55	12022	CTGAAATGTATTTTACC--ATAGACAAATCATGTGCGATAAACAAATAAGTCAATCCTTG
	12023	-----
	12024	TTACGTGCTTCATTAAGAAGGTGACGAACCTTCATCATCAATAAGTTGTGCAGTTTGAGCA
	12019	TTTTTGAAGAGCTGTTGTTGATGTAT-ATTT-ATCAGAAGCTTCATCATAGTTTCTGAT
	12021	TTATCCAAAACATTTGGTTTAATGCCTCATATGACCATTTACGAAAATATAGTTCT-TGTA
	12020	TTGTGCTAG--TTTCTATTGTTGCAT-CTAAATCCATCTCATAGATGATATTTTC---A
60	12018	TGACCCATACATATCTTGTGAAACATCTTCTGCTCGTAAAGTAATAGAAATCTATTAAG--
	12022	TTCTCTTTGCAATTTTGTGCAATTAATAAACAACCTTGTGCTTGGATTGAAACATCTAAGGC
65	12023	-----
	12024	GAATATGATTTTTTCAGGTGACATTTGACCA--GCCATCATTGCGTGGTTGCCTTCGTATT
	12019	GCAATTTTCATTACGTACATAAGCAGCAGCT--TGACGACGAGCATGTAAATCACCACGTT
	12021	CCAAAATTATTGAAATGGTCAGAGAAGCT--AAAAGA-GCTAAAGCAAGGGAACCTTATT
	12020	ACGTATTTAGTCACCTGAGCAGCTGCTACT--TCAATATTAGGAAGTAGGTCATTTTTT
	12018	---ATTGTTACTTCAACTTTTTCGGTCTT--CTCTCTGTATACTTTGAGGTTGACTCT
65	12022	AGATATTGGTTTCATCAGCAATGATAAATTTAGGCTCTACTGCTAAAGCAGTGCAATCCC

-2970-

5	12023	-----
	12024	GAACTGGTCCAAGTTTCTCGCTCATACCATATTAGTTACCATAGCGCGGGCCATAGCAG
	12019	TACCTAGAGTAATCATTTTTCAACTGTTTTACGGATTCTTTAGCACGTGCTTCAGTAG
	12021	AAATTAGTTGAATTACCCGAAGAA-TATTGGATCGCTACCCCTAGTGAGTTGTCTGGCCG
	12020	CAATAGCTTCACTTGTAGTTACAACGTTTTTATCATCAATTTTTGGAAGT--TCTGGTTG
	12018	AGTATCTAATTCTTGTGCTTC----ATTAAAGTATTTTTCAACTTTAGAGAGTTTGGTC
	12022	GATACGTTGTCTGTGTCCACCTGAAAATTATGCGGATAACGTGTTAAATGATCTTTATT
10	12023	-----
	12024	TGGCTTGTTCGAAGTCAITTTGAGGCACCTGTTGTCTGAGCGTTGAAAATAATTTCTCCG
	12019	TTACAATTGATTTCGTTGATAAGAAGATCGGTTGTCAAATCA--CGAAGCATTGCCTTACG
	12021	TCAGCAACAACGTATCGGTGTCACTTCGCGCTCTTGCAGCAGACCAAGATATTATTTTAAT
	12020	T----GATACGCTCTTCTTTTTCAAGCGT-TTCATCAACCTCCTCTATATATTCTTCCACC
	12018	TCAACATACTCA--CGAATAGCTTCTG----TTACTTCGATGTTTTACCACGAAT-ACT
15	12022	TAACCCTACAAGATCTAATAGGGCTGAACTTTACTATCACGATCTGATTTTGATTTAGC
20	12023	-----
	12024	CTACACGTCCTCCATAAGACCTGCTAATTGCTCTTTCATATCATCTTTTGAAAGAAGCA
	12019	TTGTGAGCTAGT-----GCGTCCTAGTTTACGGTAAGCCATTATGTCTCCTATTTTA
	12021	GGATGAGCCTTTT-----GGAGCTCTGGATCCTATTACTAGAGAAGGTATTCAAGACTTA
	12020	ACATCTACGCTA-----GACGGTACATTCTTAATATTTTTAACG--CTACCGATTCA
	12018	GTATTTAATCAT---ATGAGTACCTCTTTCTTGCCTGTTTAACGCTTTCTATACTCTTA
	12022	TAATTTATGTAT---ATCTAAACCT-TCTGCTACGATATCACGAATCTTCATACGGCCG
25	12023	-----
	12024	TTTGATCTTCTTTAGGT---AAAGCAATCATATATCCACCTGCACGACCACGTGGTACG
	12019	TTTATCGTTTTTTAATC---CAAGACCTAGATCGGCAAGTTTGATTTTAACTTCTTCAA
	12021	GTCAAGTCTCTTTCAGG-----AAGAAATGGG--GAAAACTATCATCTTAGTTACT-CAT
	12020	TTAATATCAGTTACTT-----CGTCGGT-GATACCTTCTATTTCAACTTTTGCTG
	12018	TTATAACC-GCTT-----TCATGAAA-----
	12022	TTTAAAGCTAGCCTGAGGATCCTGAAAAATCATCTGAGCGCTCTTTACGAAAACATGTAAT
30	12023	-----
	12024	ATAGTAACTTTATGAACAACTCGCGCATTTGAAAGAATCAAACCGACAATTGTGTGCCCA
	12019	GACTCTTACGTCCTAAGTTTCGGACTTTCATCATTTTCAGGCTCAGTTTTTTCTG-TTAAA
	12021	GA---T-ATGGATGAAGCCCTCAAGTT--AGCAACAAAATT--ATTGTTATGG-ACAAT
	12020	GC---TTTTTACCAAAGCCCAAAAAACCTTTTTTCTCACGTGATACAACTTTTATATGTG
	12018	-----
	12022	GCTTTACCTTTTCAATGTGAGATCACTTCTCCATTAAAGGTAATTTCTCCATCAGAAAATA
35	12023	-----
	12024	GCTTCATGGTAAGCAACCATAGCTCTTTCTCTTTTCAAGAAATAGTACGATCTTTTTTAGAA
	12019	TCAAATACGTATTAATTCCAGCAGCTTTTAAACAGTTATATGAGCGCACTGACAAATCA
	12021	GGTAAATGGTCCAAGAAGGGACACCCAATGATCTCTTACATCATCCTGCTA-----
	12020	CCCTCAATCGTGAAATGTTTAACTCTTGTAGTCCTTTTTCAATAGCTTCTTCTACAGTCG
	12018	-----
	12022	TCATAAAGTTTAAAAATTGAACGTCCAACGGTTGTCTTCTGATCCAGATTCCCCAACT
40	12023	-----
	12024	GGACCAGCAATTACACGGTCTTCTGCTTCATCAATATCTGAAGCATCAATAACTTTTTTA
	12019	AGTTCCCTCAATTGTCGGGTCAAGCACTTCTCATCGTTCACTTCTCTGTTTCCTTCATT
	12021	-----
	12020	CTCCTGTAAATAATACC-----
	12018	-----
	12022	AATCCAAACACTTCACTTTCATAAATGTCAAACTAACATTATCAATTGCTCTCACTTCA
45	12023	-----
	12024	TTTCGTGCGCAGCAACTAAAGCAGCTTCATTGAGAACATTTCTCCAAATCAGCACCAACA
	12019	ACTTCAGTTGCTTTAGCAACCTCTGTTAAATCAGTAAACAAGTTTAAAGTGTTCATTAAG
	12021	-----
	12020	-----
	12018	-----
	12022	TTAGCTTTTCTTTTATTGAAGGTCAAAGAAACATTTTTGACTTCAACTAATTTTTTCGA
50	12023	-----
	12024	AATCCTGGGGTTTGTGAGCTACTACTTTTAAAGTCAACATTATCTGCTAATGGTTTATTT...
	12019	ACGCGAGCTGAAAGACCAAGAGCATCCTCAGGAATGA-----
	12021	-----
	12020	-----
	12018	-----
	12022	-----
55	12023	-----
	12024	-----
	12019	-----
	12021	-----
	12020	-----
	12018	-----
	12022	-----
60	12023	-----
	12024	-----
	12019	-----
	12021	-----
	12020	-----
	12018	-----
	12022	-----
65	12023	-----
	12024	-----
	12019	-----
	12021	-----
	12020	-----
	12018	-----
	12022	-----

-2971-

12020	-----
12018	-----
12022	TTTTCAGTCATTAGGCT-----

- 5 It will be understood that the invention has been described by way of example only and modifications may be made whilst remaining within the scope and spirit of the invention.

TABLE I – THEROETICAL MOLECULAR WEIGHTS FOR GBS PROTEINS

GBS #	expected mol. weight (dalton)		
	GST-fusion	His-fusion	Native
1	78425	53460	49720
2	40035	15070	11330
3	90305	65340	61600
4	43115	18150	14410
5	158835	133870	130130
6	39265	14300	10560
7	44985	20020	16280
8	56315	31350	27610
9	50265	25300	21560
10	96465	71500	67760
11	91515	66550	62810
11d	85905	60940	57200
12	64455	39490	35750
13	40475	15510	11770
14	33325	8360	4620
15	44765	19800	16060
16	73475	48510	44770
17	46745	21780	18040
18	54335	29370	25630
19	46085	21120	17380
20	47625	22660	18920
21	56535	31570	27830
21 long	66435	41470	37730
22	60055	35090	31350
23	60165	35200	31460
24	58405	33440	29700
25	50265	25300	21560
26	118245	93280	89540
28	63795	38830	35090
29	50595	25630	21890
30	44215	19250	15510
31	63795	38830	35090
31d	58735	33770	30030
32	40585	15620	11880
33	71495	46530	42790
34	69295	44330	40590
35	56535	31570	27830
36	59065	34100	30360
37	46965	22000	18260
38	61815	36850	33110
39	65225	40260	36520
41	75235	50270	46530
42	46745	21780	18040
43	58955	33990	30250
44	52355	27390	23650
45	43555	18590	14850
46	59835	34870	31130
47	84255	59290	55550
48	86455	61490	57750
48d	106695	81730	77990
49	59615	34650	30910
50	94155	69190	65450

51	47075	22110	18370
52	55435	30470	26730
53	110215	85250	81510
54	73365	48400	44660
55	36295	11330	7590
56	34865	9900	6160
57	51145	26180	22440
58	128805	103840	100100
59	99215	74250	70510
60	63575	38610	34870
61	68085	43120	39380
62	105485	80520	76780
63	64125	39160	35420
64	112745	87780	84040
65	72485	47520	43780
66	49715	24750	21010
67	120335	95370	91630
68	131225	106260	102520
68d	103065	78100	74360
69	53895	28930	25190
70	74465	49500	45760
70d	59725	34760	31020
71	56755	31790	28050
72	75565	50600	46860
73	72815	47850	44110
74	131225	106260	102520
74d	95475	70510	66770
75	114725	89760	86020
76	198875	173910	170170
77	78535	53570	49830
78	48835	23870	20130
79	58185	33220	29480
79d	50815	25850	22110
80	81835	56870	53130
81	89205	64240	60500
82	40475	15510	11770
83	62585	37620	33880
84	122645	97680	93940
85	70175	45210	41470
86	84035	59070	55330
87	44435	19470	15730
88	73365	48400	44660
89	143325	118360	114620
90	93495	68530	64790
91	88325	63360	59620
92	193595	168630	164890
93	95585	70620	66880
94	77435	52470	48730
95	60605	35640	31900
96	57195	32230	28490
97	138375	113410	109670
98	82055	57090	53350
99	60715	35750	32010
100	53015	28050	24310
101	59395	34430	30690
102	40695	15730	11990
103	56975	32010	28270

-2974-

104	120005	95040	91300
105	179735	154770	151030
105dNterm	127265	102300	98560
105dCterm	81285	56320	52580
106	85795	60830	57090
107	89535	64570	60830
108	64565	39600	35860
109	75125	50160	46420
109d	70725	45760	42020
110	53895	28930	25190
111 / 190	60165	35200	31460
112	63905	38940	35200
113	59175	34210	30470
114	51915	26950	23210
115	98225	73260	69520
116	73475	48510	44770
117	47515	22550	18810
118	42235	17270	13530
119	109225	84260	80520
120	71385	46420	42680
121	65115	40150	36410
122	46855	21890	18150
123	68305	43340	39600
124	54115	29150	25410
125	57305	32340	28600
126	56865	31900	28160
127	80845	55880	52140
128	39925	14960	11220
129	43775	18810	15070
130	82275	57310	53570
130d	63245	38280	34540
131	89755	64790	61050
132	49055	24090	20350
133	54445	29480	25740
134	42015	17050	13310
135	65225	40260	36520
136	54885	29920	26180
137	63465	38500	34760
138	40145	15180	11440
139	38165	13200	9460
140	43445	18480	14740
141	49935	24970	21230
142	79745	54780	51040
143	33545	8580	4840
144	49165	24200	20460
145	63025	38060	34320
146	107025	82060	78320
147	156965	132000	128260
148	41905	16940	13200
149	62365	37400	33660
150	54665	29700	25960
151	50412	25447	21707
151L	50045	25080	21340
152	45535	20570	16830
153	46965	22000	18260
154	101525	76560	72820
155	62585	37620	33880

-2975-

156	61265	36300	32560
157	74025	49060	45320
158	52025	27060	23320
159	41025	16060	12320
160	82825	57860	54120
161	95365	70400	66660
162	42015	17050	13310
163	69405	44440	40700
164	42345	17380	13640
165	43555	18590	14850
166	38055	13090	9350
167	50375	25410	21670
168	32555	7590	3850
169	43445	18480	14740
170	64015	39050	35310
170d	59945	34980	31240
171	49825	24860	21120
172	62365	37400	33660
173	96795	71830	68090
174	45095	20130	16390
175	59175	34210	30470
176	55435	30470	26730
177	66215	41250	37510
178	62365	37400	33660
179	58515	33550	29810
180	37615	12650	8910
181	63685	38720	34980
182	90085	65120	61380
182d	87225	62260	58520
183	57855	32890	29150
184	46415	21450	17710
185	40695	15730	11990
186	85685	60720	56980
187	56205	31240	27500
188	61595	36630	32890
189	60165	35200	31460
191	116705	91740	88000
192	69625	44660	40920
193	98005	73040	69300
194	49385	24420	20680
195	81065	56100	52360
195L	147615	122650	118910
195L N-term	91405	66440	62700
196	69515	44550	40810
197	99325	74360	70620
198	73805	48840	45100
199	158285	133320	129580
200	132325	107360	103620
201	74538	49573	45833
202	157295	132330	128590
203	61705	36740	33000
204	39705	14740	11000
205	55985	31020	27280
206	56645	31680	27940
207	44765	19800	16060
208	59725	34760	31020
209	62145	37180	33440

-2976-

209d	56425	31460	27720
210	60935	35970	32230
210d	53675	28710	24970
211	64895	39930	36190
212	60825	35860	32120
213	45205	20240	16500
214	38935	13970	10230
215	45205	20240	16500
216	91515	66550	62810
217	36075	11110	7370
218	81065	56100	52360
219	56535	31570	27830
220	54555	29590	25850
220	50155	25190	21450
221	41465	16500	12760
222	47405	22440	18700
223	42895	17930	14190
224	45865	20900	17160
225	56645	31680	27940
226	44875	19910	16170
227	46195	21230	17490
228	46525	21560	17820
229	35855	10890	7150
230	51915	26950	23210
231	60935	35970	32230
231d	58735	33770	30030
232	41795	16830	13090
233	35635	10670	6930
234	43115	18150	14410
235	58295	33330	29590
235d	48395	23430	19690
236	46525	21560	17820
237	44215	19250	15510
238	59725	34760	31020
239	63905	38940	35200
240	51475	26510	22770
241	45095	20130	16390
242	43225	18260	14520
243	119455	94490	90750
244	48065	23100	19360
245	48615	23650	19910
246	49605	24640	20900
246d	45975	21010	17270
247	58955	33990	30250
248	92505	67540	63800
248d	70835	45870	42130
249	103835	78870	75130
250	136505	111540	107800
251	52135	27170	23430
252	51695	26730	22990
253	74245	49280	45540
254	59615	34650	30910
255	69075	44110	40370
256	47845	22880	19140
257	60495	35530	31790
258	67975	43010	39270
259	79415	54450	50710

-2977-

260	48175	23210	19470
261	55765	30800	27060
262	75345	50380	46640
263	63465	38500	34760
264	47185	22220	18480
265	56315	31350	27610
266	51365	26400	22660
267	88655	63690	59950
268	50265	25300	21560
269	60495	35530	31790
270	59285	34320	30580
271	56315	31350	27610
272	118355	93390	89650
272d	98885	73920	70180
273	70945	45980	42240
274	56205	31240	27500
275	47515	22550	18810
276	147945	122980	119240
277	87005	62040	58300
277d	75675	50710	46970
278	52245	27280	23540
279	79415	54450	50710
280	88655	63690	59950
281	74465	49500	45760
281d	71495	46530	42790
282	44765	19800	16060
283		20240	16500
284	67645	42680	38940
285	57525	32560	28820
286	41355	16390	12650
287	61045	36080	32340
287d	57085	32120	28380
288	53675	28710	24970
288d	51035	26070	22330
289	65005	40040	36300
289 long	71825	46860	43120
290	47405	22440	18700
291	63795	38830	35090
292	103505	78540	74800
293	115935	90970	87230
293d N-term	73805	48840	45100
293d C-term	70835	45870	42130
294	75785	50820	47080
295	89425	64460	60720
296	60385	35420	31680
297	100205	75240	71500
298	54335	29370	25630
299	62255	37290	33550
300	130895	105930	102190
301	54885	29920	26180
302	80075	55110	51370
303	53235	28270	24530
304	75125	50160	46420
305	78645	53680	49940
306	67975	43010	39270
307	86675	61710	57970
308	59285	34320	30580

-2978-

309	62695	37730	33990
310	58845	33880	30140
311	76445	51480	47740
312	64785	39820	36080
313	65995	41030	37290
314	52135	27170	23430
315	51695	26730	22990
316	41795	16830	13090
317	179295	154330	150590
317d N-term	115935	90970	87230
317d C-term	92160	67402	63360
318	70065	45100	41360
319	61925	36960	33220
320	57965	33000	29260
321	83705	58740	55000
322	76628	51663	47923
323	86345	61380	57640
324	86345	61380	57640
325	82605	57640	53900
326	91515	66550	62810
326L	172695	147730	143990
326L N-term	113955	88990	85250
327	279175	254210	250470
327d N-term	139915	114950	111210
327d C-term	167965	143000	139260
328	97602	72637	68897
329	113955	88990	85250
330	83595	58630	54890
331	60825	35860	32120
332	75675	50710	46970
333	63465	38500	34760
333d	57965	33000	29260
334	38275	13310	9570
335	43555	18590	14850
336	67645	42680	38940
337	75235	50270	46530
338	54995	30030	26290
339	76665	51700	47960
339d	72925	47960	44220
340	86565	61600	57860
341	38385	13420	9680
342	61595	36630	32890
343	60385	35420	31680
344	55875	30910	27170
345	40585	15620	11880
346	53895	28930	25190
347	55325	30360	26620
348	58405	33440	29700
349	98335	73370	69630
350	53895	28930	25190
351	82165	57200	53460
352	111315	86350	82610
352d	105485	80520	76780
353	55325	30360	26620
354	42345	17380	13640
355	52135	27170	23430
356	59065	34100	30360

-2979-

357	40255	15290	11550
358	60495	35530	31790
359	78865	53900	50160
360	73695	48730	44990
361	109005	84040	80300
362	125945	100980	97240
362d N-term	63355	38390	34650
362d C-term	91295	66330	62590
363	53125	28160	24420
364	75015	50050	46310
365	102075	77110	73370
366	68415	43450	39710
367	76885	51920	48180
368	44765	19800	16060
369	142115	117150	113410
370	94595	69630	65890
371	65555	40590	36850
372	55105	30140	26400
373	50265	25300	21560
374	57525	32560	28820
375	66875	41910	38170
376	48065	23100	19360
377	73805	48840	45100
378	58955	33990	30250
379	68855	43890	40150
380	47405	22440	18700
381	66875	41910	38170
382	50815	25850	22110
383	57085	32120	28380
384	77985	53020	49280
385	75675	50710	46970
386	39485	14520	10780
387	54555	29590	25850
388	45645	20680	16940
389	43005	18040	14300
390	62255	37290	33550
391	54775	29810	26070
392	71385	46420	42680
393	55765	30800	27060
394	59725	34760	31020
395	72375	47410	43670
396	34865	9900	6160
397	113625	88660	84920
397d	100865	3740	72160
398	56755	31790	28050
399	55435	30470	26730
400	74135	49170	45430
401	59395	34430	30690
402	78095	53130	49390
403	64455	39490	35750
404	61595	36630	32890
405	45975	21010	17270
406	36955	11990	8250
407	82715	57750	54010
407d	71715	46750	43010
408	45315	20350	16610
409	70395	45430	41690

-2980-

409d	59600	34842	30800
410	62475	37510	33770
411	41355	16390	12650
412	35965	11000	7260
413	59175	34210	30470
414	50375	25410	21670
415	46195	21230	17490
416	42455	17490	13750
417	77985	53020	49280
418	42125	17160	13420
419	47515	22550	18810
420	67755	42790	39050
421	62915	37950	34210
422	60165	35200	31460
423	74245	49280	45540
424	89975	65010	61270
424	77325	52360	48620
425	116045	91080	87340
426	83815	58850	55110
427	41135	16170	12430
428	55325	30360	26620
429	59175	34210	30470
430	53785	28820	25080
431	54005	29040	25300
432	65665	40700	36960
433	40915	15950	12210
434	44545	19580	15840
642	91845	66880	63140
643	78975	54010	50270
644	49605	24640	20900
645	59725	34760	31020
646	61595	36630	32890
647	55875	30910	27170
648	59835	34870	31130
649	76115	51150	47410
650	51475	26510	22770
651	53345	28380	24640
652	49715	24750	21010
653	44655	19690	15950
654	51255	26290	22550
655	65995	41030	37290
656	57525	32560	28820
657	62805	37840	34100
658	60165	35200	31460
659	60275	35310	31570
660	71495	46530	42790
661	60605	35640	31900
662	62695	37730	33990
663	89535	64570	60830
664	45315	20350	16610
665	41135	16170	12430
666	47075	22110	18370
667	53162	28197	24457
668	43555	18590	14850
669	48505	23540	19800
670	45315	20350	16610
671	36940	12182	8140

-2981-

672	40130	15372	11330
673	41450	16692	12650
674	45300	20542	16500
675	55970	31212	27170
676	65650	40892	36850
677	54320	29562	25520
678	77750	52992	48950
679	60480	35722	31680
680	64440	39682	35640
681	93040	68282	64240
682	84790	60032	55990
683	15950	44655	19690
684	11880	40585	15620
685	16280	44985	20020
686	21340	50045	25080
687	9350	38055	13090
689	55105	3740	26400

TABLE II – PRIMERS USED TO AMPLIFY GBS_{nnn} PROTEINS

Forward primers begin 5' -GGGGACAAGTTTGTACAAAAAAGCAGGC-3' and continue with the sequences indicated in the table below; reverse primers begin 5' -GGGGACCACTTTGTACAAGAAAGCTGGGTT-3' and continue with the sequences indicated in the table. The primers for GBS1 are thus:

Fwd: GGGGACAAGTTTGTACAAAAAAGCAGGC**TCTCAATCTCATATTGTTTCAG**

Rev: GGGGACCACTTTGTACAAGAAAGCTGGGTT**ATTTTTAGACATCATAGACA**

The full forward primer sequences are given in the sequence listing as SEQ IDs 10968-11492. The reverse primer sequences are SEQ IDs 11493-12017.

GBS	Forward	Reverse
1	TCTCAATCTCATATTGTTTCAG	ATTTTGTAGACATCATAGACA
2	TCTAATTACATTATTACATTTTG	GGGAATGCCTACAAA
3	TCTGATACTAGTTCAGGAATATC	TTTTTACTATACTTTTTGT
4	TCTGATACAAGTGATAAGAATACT	TTCCTTTTTAGGCTTACT
5	TCTATTTTTCTTCATAGTCCAC	ATTAGCTTCATTTGTCAG
6	TCTGAATGGGTGTTATTAATCTC	AGTTTCTTCTTAAATCAT
7	TCTACAAATTCTTATTTTAGCAA	CTCTGAAGCTGTAAACC
8	TCTGTATCAGTTCAGGCGT	TTTATCAATGTTTGAACG
9	TCTGCTGCTCTAGGACAAC	TAGTAAATCAAGTTTTTGCA
10	TCTTTTGTTGTTGCCTTATT	ATCCCTTCTATTTTCGA
11	TCTCCACCTATGGAACGT	ATGTAGTGACGTTTCTGTG
11d	TCTCAGAAAGTCTATCGGG	ATGTAGTGACGTTTCTGTG
12	TCTAGTGAGAAGAAAGCAAAT	ATTGGGTGTAAGCATT
13	TCTTCTTGGAATTATTGGAG	CTTAACCTACCCGTCC
14	TCTGCAATGATTGTAACCAT	TTTTCTCTTATTAAAGAATT
15	TCTGCATCTTATACCGTGAA	ATACCAGCCGTTACTATT
16	TCTGCCGAGAAGGATAAA	TTTAGCTGCTTTTTTAATG
17	TCTGTTTATAAAGTTATTCAAAA	AAATACTACATTTACAGGTG
18	TCTAAGCCTAACAGTCAACA	TTGGTTATTCTCCTTTAAT
19	TCTGATGATAACTTTGAAATGC	ATTATATTTTTGGATATTC
20	TCTGCAGTGATTGCAAGTC	GGGCTTTTTCTTAAAAA
21	TGTGCTGCATCAAAC	GTTGGCATCCCTTTT
21 Long+A527	TGTGCTGCATCAAAC	CTTTTGATGGGATTGG

22	TGTAATAACAAAGCCAG	TTGATTTAACGATTTGA
23	TGTCAATTAACCGATAC	TTTATCTCCTCTAAAATAATG
24	TGCTCAAATGATTCAT	CTTTGATAAGTCAGACCA
25	TCTAAAAGTTCACAAGTTACTACT	GTAACCCCAAGCTGAT
26	TCTAGTCATTATTCCATAAAATT	TGATTTTGCAATATCAA
28	TCTAATCATATGCTGATTGAG	TTTTTGTAATTTAAGTACTAA
29	TCAGTTTGGATGTTAAC	TTCTTTTATATTAAGAGCTT
30	TCAACAAATGCAGATG	ATTCGGATAAATGTAGC
31	TGTTTTGTCATTATTGATAG	TCCATTTTATCCTCAC
31d	TCTCTAACTTGGTTTTATTAGA	TCCATTTTATCCTCAC
32	TCTGGTTTAAAAGTGACTGAA	ATGACCTCTACTTTCCA
33	TCTCATCATTTAGGTAAGGAA	CTTGTAATCACTTGGAC
34	TCTGTTAGTAATCGCTACAATC	ATTAATCATGGTATTGGT
35	TCTAATCAAGAAGTTTCAGC	CCATTGTGGAATATCA
36	TCTCGAGTTTTAGCGGATA	TTTGTAAGCAGTTCTT
37	TCTGTATTATTTACCAATCACA	ATCATTATCATGATCTCTAGA
38	TTAGGAGTGGTAGTTCAT	ATTTTGATTGATTCTACTC
39	TTTTTATTGTTAGTATTAGC	TTTTGTTTTTTCAAATA
41	TCTGTTTATCTAGCGGTTAGA	ATCTTCAACGTCTCTCC
42	TATAACAGTTTAGTTAGAAGTC	AAAGTCAAAGGAACTT
43	TTTAAAGGGTTTACATATT	TTCTTTATCTAATTTATAATAG
44	TTTAATACAATTGGTCG	TTGCAATGTTTTTCT
45	TCTATGGAAAAAATTAGGATT	TAACTTTGGATAATCTGT
46	TCTAGAGATGAGCAAGAAATA	GTTGAAATTTTGATATGA
47	TCTCAACAGATAGGTCCTTTATAA	CTCCTTTACTATATAGCTAACT
48	TTTCTCTATAATTACTTCAAT	TTGTTTGTGAAGTAAAC
49	TCTAATAAGGCATTATTAGAGG	TGATAATATCTCCATATTTT
50	TCTACACATTTAGTTGACTTAAC	GCATTGGCGCCATA
51	TCTAGTAAACAACACATTTATCTA	TTCTACACGACTTTTATTC
52	TCTCAAGAACTCATCAGTTG	AAGACCTCCTCGAGAT
53	TCTGCAGAAGACATTGTTACA	TGTTTTTCTTTCTGTTG
54	TATAATTTTTCGACTAATGA	TGGATTAGTTTGACCTG
55	TCTGACACAGTGTCTTATCCT	TTTATCGTAAGCACTTAGG
56	TCTGTGGAGCAAGTGGCCA	CTCCTTCCAGGCATCG
57	TCTCAAGAACTAAGTAACTTTGA	GTAAAAGTATCTTAAATAGTCA
58	TCTACTGAAACGTTTGAAGG	TGCCATTCCTCTCT
59	TCTGATGAAGCAACAATAA	TGTTACCTTTTTATTTCT
60	TCTAATAAAGATAATCAAAAACT	TTTTTCATGCGATTGA
61	TGTTTCTTTTTTATTCCA	GAGACGTTTCTTATACCTT
62	TATTACTTTGATGGTAGTTT	TGTACCATATGTTCTCTCT
63	TCTGTTCAATCATTAGCAAA	AAAAGTTGGACTACTTTC
64	TTTAAAGGTAATAAGAAGTTG	TCGTTTTCCACCC
64d	TCTAGTCAAGTTGACTCTGTTA	TCGTTTTCCACCC
65	TCTCAAAACCAGGTGACTG	ATTTGGGTAAATATAGTAA
66	TTAAGATTTTATAACAACGA	TTTACGACTAACCTCAAC
67	TCTAATGTTTTAGGGGAAA	AATTCCTTTTGGTGG
68	TCCCAAAAGACTTTTG	GGCAGAATACACCTTC
68d	TCCCAAAAGACTTTTG	GGCTGACGTCGACGCA
69	TCTAAAGTTTTAGCCTTTGA	AACTCTCTTAATATATTCTTCT
70	TCTGAAATGGCTTTAG	GTCTTTTTCAATATTCTGT
70d	TCTACTAACTTATTGAGTAGAATCA	GTCTTTTTCAATATTCTGT
71	TGTAGCTCAAAATCTCAT	CTTCTCCTTAGGAGTAACG
72	TCTAGTTTATCTATTAAGATGCC	ATTATTATCAATTAATAACTCTT
73	TCTATCAAAGAGGCGGTAA	GTCAAACATACTTCCAAA
74	TCTAAAGAGGATAAAAAGCTAG	TTTCGTCGTATAAGCA
74d	TCTAGTGTTCAGGTAGTAGTG	TTTCGTCGTATAAGCA
75	TCTAAAAAATTAACACTCAA	TGTCCTCATTTTTTCAG

76	TCTGATGAAGTTACAACTTCAG	AATACTTGCTGGAACAG
77	TTATTCCAAAGTAAAAATAAA	GTCTTTCTTCAATTTTGG
78	TCTCATAACCATCACTCAGAACACATGT	GTCGTGATTTTTATGAGT
79	TCTCCCAAGAATAGGATAAA	CCCAAAGTGGCATAAC
79d	TCTAGTCAGTATGAGTCACAGA	CCCAAAGTGGCATAAC
80	TCTGCAGAAGTGTCACAAGA	TGAAGGACGTTTGTG
81	TCTTTTGATGGATTTTT	TTTTTTAGTTAAGGCTA
82	TCTACAAATGAAAAACGAAC	GTCCACCTTCCGAT
83	TCTGAAATTAAGTCAAAAAATATT	AACATTGTTTTCTTTT
84	TCTCATACTCAAGAACACAAAA	ATGGTGATGATGACCT
85	TCTCCTAAGAAGAAATCAGATAC	ATTAACATTTTGAGGT
86	TCTGCAGAACTAACTCTTTTAA	TTTGC AAAATCAACA
87	TCTGCGGATACATATAATAACTA	GAATAAATAACTGTATTTTT
88	TCTTACCAAAAAATGACG	ATTTTCATTAATTTCTCT
89	TCTGAAGAGCTTACCAAAAC	GATAGCTAATTGGTCTGT
90	TCTAGATATACAAATGGAAATTT	TAAAAGATGAGCTTCTCG
91	TCTAAAAAAGGACAAGTAAATG	AATTTCAATATAGCGACG
92	TCTGATTCTGTCATAAATAAGC	CTTGTTGTCTTTACCTT
93	TCTGAATTTTCACGAGAAA	ATTATCCTTCAAAGCTG
94	TACCAATTAGGTAGCTATAA	TGTGTCATATAATGTAACCA
95	TCTGTTAATACAAAAACACTTCT	TGATCTTAATTTTCGAG
96	TCTGGTCAGTCTAAAAATGAAG	CCAAACAGGTTGATCT
97	TCTAGCCAGGAGGTATATG	ATTTACATCAGACTGTGAC
98	TCTGAACTATTAATCCAGAAA	TTTATGGCCAATAACA
99	TCTACAAGTATGAACCATCAA	TTTTTTAGTAGTTGTCAATT
100	TCTAAGGGGCCAAAAGTAG	GTAAGCTGAATTTTCGA
101	TCTATTACTTTAGAAAAATTTATAGA	ACGAGAGTGGTTATTGG
102	TCTGCCTTTTACTTTGGCA	TTTCTTCACTCTTTCTAGAG
103	TCTATTTTTTCTTGATCAT	CGGCCAGTTTTTCTT
104	TCTGGTGAAACCCAAGATA	AACACCTGGTGGGCGT
105	TTAACAATTCATGGACC	ACTATTTCTAATTGCTCTG
105d	TTAACAATTCATGGACC	TGGTCCCGGTGCGCCA
105d	TCTCAAGGACCTCCCGGTG	ACTATTTCTAATTGCTCTG
106	TCTCAAAATCAAAATTCACA	CTTAGCAGATTCATCCC
107	TCTCTGGAGCCTTTTATTT	TTTACTATTTGAAAATTGG
108	TCTGGTAATCGTTCAGATAAG	TTTCATAGGAAGTTGTATT
109	TCTATCCAGCAGATCAACT	GTCCACACCTGCGACT
109d	TCTAAACGGGTTTCGCTATG	GTCCACACCTGCGACT
110	TCTGTAAAATTAGTATTCGCAC	TTTACCTAAGTAATATTCTGA
111.19	TCTGTTAGCGTTGATAAGGC	TCCCGTCTTTTTTGT
112	TCTACAATTA AAAATCTCACTG	GTCGTAATCATAAAAGCC
113	TCTAGTAAAATCAAAATTGTAACG	TTCATAACGAACCATAAC
114	TCTAATCTTTTAATTATGGGTT	TTTGAGTTCTAGCAACG
115	TTTCAATACTATTTAAAAGG	TTTTTTATCTTCTTCTG
116	TCTACCGAGGAGCCATTAA	TTTTAAAACCTGGTAAAC
117	TCTGAACAATCACAAAAACA	TCAGCTCGTACTGTTT
118	TCTATGGTGACGGTGCTGG	GTCCTCCTCAATTGGT
119	TCTAGTCAGCCGGTAGGGG	CTCTTTTATACGCGATG
120	TCTGGTGGAGCATTGCTA	GTTATTTGCTCGTTGTT
121	TCTAATAAAGATAATCAAAAACT	TTTCTCAAATGTTTTCAT
122	TCTGCTGCCACCAAGAAAG	TTTCAAATGATCTACAGC
123	TCTACAACAAATGTAATGGC	GGCTAGTGTCTGTCCG
124	TCAATGAATTTTTCATTT	ACCATCTATTTTACCCC
125	TCTACAAAATATCAGCGAATG	AGAACCCGCACTCTCA
126	TCTACTAAGCAAGCAATGTC	GAACGCAACGGCTGCT
127	TCTACAAAAGAATATCAAAATTAT	TTTCATATCAAAAATATCG
128	TCGACTAATTCGTTAAA	TTCTTTATCTCTTAATGCTT

129	TTTGAAATAGTATTGGAAA	CACAACAGTTATTTTTTCA
130	TCTATATTTCTATTTTATTATGT	AGGCCCTTCTGAGTAG
130d	TCTAAAAACAACCTTCACAAC	AGGCCCTTCTGAGTAG
131	TCTAAACAGATATTGAAATAGC	AAATAATCCAATGGCTG
132	TCTATTAATATTATCATTTGCA	CTTTTCAAGCTTTTCC
133	TCTGCTTTACGGAACCTTG	AAAATGATCAGTTTGAGG
134	TCTACTATTTCTCAACAACAATAC	TTTTTGGCTTAAGAAAG
135	TCTGAAAAAAGAGTAGTTCAAC	CTTACGATACATTTTAAATTG
136	TCTAATCAATTATCAGAAATCA	TTCTTTTTTACTTTAGCG
137	TCTCAAGAGTATAAAACAAAAGAG	CCATTGCAATCCAGCA
138	TCTGCTGTATTTACACTCGTC	ATGTTTATGGCTTGCT
139	TCTGGCGGCAAGATAAAAT	TTTTTGATAAATCCCC
140	TCTGATGGGTAAAGAATAATG	ATATGTGTATTCATCCTTT
141	TCTGATGTTGTAATTAGTGGAG	TACTTCTATTTTCCATCTG
142	TTCGAATTAAGAGAAAAGA	GTAATGCAATAAATCAAAA
143	TCTAGCTTTTATGTATTTCA	GGATTTTAGTTTCGCA
144	TATACGCATAGTGGAAC	CCCATTGATTCGTCTG
145	TCTGTTATTATCAGGGGCG	TACCTCTTCAATACCAC
146	TCTGTTAGTCGTTCTCCGA	ATTACCGTTAGGTACTGTA
147	TCTGAGGAGCAAGAATTAAG	GGTATGGTTAACAGAATC
148	TCTATTCTAACAAAAGCAAGT	ATATACCCTAGACTTTTTGA
149	TCTAGTGGGCGTTTCATGGA	AGGAGTTTTATTGATGATAT
150	TCTGATACCCCTAATCAACTA	AAATGATTGTGGAAAAA
151	TGCAGGAGCTGTCCGC	ATCAAAGAAGTTGACATTG
151 Long	TCTGTCCGCATTGGTAAAG	ATCAAAGAAGTTGACATTG
152	TCTAACTGCTTAGAAAATGAA	GTTAGATAAATTAACCAGTG
153	TCTAACTGCTTAGAAAATGAA	CCCTTTGCTTCGTTGT
154	TCTGGAAAGGTCAGTGCAG	TTCCACAAGTCCGATT
155	TCTATTTTATTTTTCAGATGAAC	TTGTTTGATTGCTCCT
156	TCTGCATCAGATGTTTCTAGA	ACTACCAAACTGCTGG
157	TCTAGTGACGTTGACAAATA	TTGTGTATTTTATGTTAGGT
158	TCTATGACCATTTACTTCAATA	GTGGATAAAATTCGAAA
159	TCTCAAACTATTTTGACGC	CAGACTGACTAGGAGCT
160	TCTGATGAATATCTACGTGTCTG	GACTTGTAATTGATTCTGC
161	TCTGATGAGGTGGACTATAACA	GAAGGCACCACCACCT
162	TCTATTTTCTTGCTCTTAGTTG	GTTGTATAGATGAGTTAATCTG
163	TCTGAACTGTCATTCAACTTG	ACGGTTTTTAAAGAATG
164	TATTTTTTAACAACAAAAAA	TTTTTCTTTATCTTCTGTG
165	TCTCCAATTTTATTGGTTT	CGATTTTGTAAGAGCTT
166	TCTGCATCTTATACCGTGAA	CGACGAAGCTATTTCT
167	TCTACAATTTATATTGCTTGG	TAAGGCTTGCATTTTG
168	TCTGTTGGATTGATGTTGG	TTTTCTAAAAATTTTCC
169	TGGAAACAAATCACAG	GGCATCTCCTAGCTTT
170	TCTGCAATAGTTTTTACTTTTTT	TGATAAAGGTAGTTCTACAC
170d	TCTGGTTCTTATCATTTAACAA	TGATAAAGGTAGTTCTACAC
171	TCTGCTAGACCCAAACAGT	TTTTAGATGTTTTTGTGG
172	TACACTCATATTGTTGAAAA	ATGATTGATAATTTTAAGC
173	TCTAATAGTACTGAGACAAGTGC	TGCTTTTTGATATGCC
174	TCTGCTTATGTCGTCAATTT	TAAATAAAGTTTCAAGAAAG
175	TCTGAATTACCTTCGTTTATC	TTTCTCCCTTGACTTTC
176	TCTAAACATCCGATACTTAATG	CTTTTTCTCAGATGCTT
177	TCTAATTATCCTTTTGCGA	GACATTGAAACGGAAT
178	TCTGGACTACGCGGAGTAT	TTTTATCAATGATGTTGA
179	TCTGCTATTGGAGCAGCTG	CATATGACGCAACGC
180	TCTGATAAAGAAGGGATAGAGG	AGCCTCTTTCTTGTT
181	TCTAAAGAAAAATCACAACTG	ACGATTATCAACAAAGTT
182	TCTCAAAATAATAAAAAAGTAAAA	CATTCTTTTAAATACAAATC

182d	TCTCAAAATAATAAAAAAGTAAAA	GGGTTTGAAAGTTTTTC
183	TCAAATGGTCAATCTAGC	TTTAACCTTTAATTACTGGAAT
184	TCTAAGGATTCAAAAATCCC	TTTTTTAATAAGCTTCGA
185	TCTGGGCAACCATCTACAT	TTTTTTGTAACTTCCTG
186	TCTCATTACAGGATAGCA	CTTAGATACATTGTTTTTTC
187	TCTGGACGAGGAGAAGTATC	CTTCTTTTCTTACTTGC
188	TCACAATCTTCTCAAAA	TTTATTATTTTAACTTGAA
189	TCTGATAAGTCAGCAAACCC	CTTCAACTGTTGATAGAGC
191	TCTATCACGACATTACAGACT	TCCTTTAGCAGGAGCT
192	TCTAGATATTTAACTGCTGGT	GTTATACATGTTGTCTGAAG
193	TCTATAAAATATCAAGATGATTTT	CCAAATAATAACACGTTT
194	TTAGAAAGTCAGAGAGCAG	GCTATCCCTTTCCAAT
195	TCTATTATGGAGACGGGTA	TGTATTTTTAATTTGTTTTC
195L	TCTTTGAATAATAAAGGTGTCTG	TGTATTTTTAATTTGTTTTC
195LN	TCTTTGAATAATAAAGGTGTCTG	CAAACCTTTAACATTTAATG
196	TCTATTTCTCAAATTTTACG	ATAGTGTAAGCTACCAGC
197	TCTAATTTTATAAGCTCTTG	GTCATCATATTCCTGAAA
198	TCTGCGCTTAAAGAATTAA	TGTTGCGCGTAAGATT
199	TTTTTAAAGAAATTGAAA	ATTGGTCATTTCTTGAG
200	TTTCGTAAATATAATTTTGA	AACAGATTTATTGGTTGG
201	TCTAGCGATACCTTTAATTTT	AGACTCATCAACTTTTTCT
202	TCTATGCTGATTAAGTCGC	GAACCCTGAAGGGTAG
203	TGTGGTAAAACTGGACT	CCAATTGTATTTTCAAC
204	TCTAAGACAGGAGCACCCGT	ATTTATACTACCTGTTGAATC
205	TGCGAGTCAATTGAGC	TTTAAATTTGTAGTCTTTAATA
206	TCTACAAATACTTTGAAAAAAGA	CTCTTTTACTTTTCCAAAA
207	TCTAATTTATTTAAACGTTCCCT	CCCTCCCTTAAGAGAA
208	TCTAAAAAGCGGCTAGTCA	TTGACGATGTTGCATC
209	TCTGGACAAAAATCAAAAATA	TTTCGAATTATTGTGACT
209d	TCTGGACAAAAATCAAAAATA	GTATTGTTGTTGCCTG
210	TCTGGAGGAAAATTTAGAA	TTTTTGATTTCCCTTTC
210d	TCTACCTCATATCCTTTTATTT	TTTATAGTGTGTTTGCAA
211	TGTGGACATCGTGGTG	TTTGCTAGGAACCTTGA
212	TCTAAGACTAAAAAATCATCA	TGATTCAATTCCTTTTC
213	TCTAAACACACCAGTAAAGAA	TTTTTCCTCTACTTTCTTA
214	TCTAAAAATAAAAAATCTTATTT	TTTGCTCACCTCCACA
215	TTAATAAAAGGATTATTGTCA	CAATAACTTCTGTAAAATAAA
216	TCTGCTCGTTTAATACCACA	TTCACCCTTAAATAAATT
217	TCTAACACTAACATCCCTAGC	TGCATTTTTCCCTTCT
218	TCTAGAGGGAAGGTTATTTAC	CTCCAGTAAAGTATTAGTATTT
219	TCTATCAATAAAGTAACAGCTCA	GTGAGGTTTTTGTAATT
220	TCTAGAACACTATTTAGAAATGATAT	TGCATATAAGTTTTTAGC
220d	TACTATGCGAATCACAG	TGCATATAAGTTTTTAGC
221	TCTAGTTTAGCATTGCAAAT	CTCATCTAAAGTGCTATCC
222	TCTACATTTTATAAAAAGACGG	CTCGTATTTAGGCAACT
223	TCTAAGAAAATACGAAGCTATAC	ATTGGATATGCCATAAA
224	TCTGGAGGAAATGAAATATTA	GACTTTTTGATGTTTACTTT
225	TCTGGTATGTCTAATAAGGAAAT	TTCTTTACTATAAACATCTTCA
226	TCTAACAACTTATTACAGAAAA	AGCATTTAAAGTTGAATGT
227	TCTGTTTCATATGAAAAAGTCC	GTTAGTCTCTTCAAGATCA
228	TCTAGTAGAGGTATTTTTTACAA	AAGACCTACCGCCCAA
229	TCTGAACGTCGGGTAAGTC	TACTTCTTTCTCTTTCAATT
230	TTTTTAATCGATTTTATTT	CTTAGTGTTCCGATATGA
231	TCATTAATTATTCTTACGGT	TCTTGTTTTAAGAGCAGA
231d	TCTTTATACGTTGTTAAACA	TCTTGTTTTAAGAGCAGA
232	TGGCTAAGTAAGCATGAG	ATCATGTTTTCCCTCAA
233	TTCCAGCTAGCTGTC	ATCTGATATATCCGTTTTAT

234	TCTATAGAAATTGCTGTATTAATT	TTTTTGTCTCCTTTTTTA
235	TCTATTCGATTTCTTATTCTTG	AAAGACACGATAAACATAAG
235d	TCTGACTCAACCACAGTCTC	AAAGACACGATAAACATAAG
236	TCTGCAGACCTTACAAGTCA	ATTTGCAACTTCTTGTATA
237	TCTATTGTATTTGCTATTGCA	TTTAAAAGTATCCTTAAATAAG
238	TCTGATATTTTTTCAGCTATTGA	CTTCCTCCTCAATAGTTG
239	TCTGTTAGTGCTGCTATTGAA	TTCTCCTCCCCATTA
240	TCTAAGAAGCTTACTTTTATTTG	ATCCAAACGAGTGAAAT
241	TCAAAAGGATATTCAAGA	AGGTGTTGTTGTATTTTC
242	TCTCATAATATATTAAGATTTTAGG	CTTTCTAAGTTTATTAAACATA
243	TCTATTCTTGGTCAAGATGT	GGCATCTGTTACCTTG
244	TCTCATGAAAATGTTAAAAAAG	AAACAACCTCCATTATTTTT
245	TCTAAGTCAACGGTAACAAA	TAAACGTTGAAGAGCAT
246	AGGAAACGTTTTTCCT	CTTATCATATCTTGTTAAATCA
246d	TCTAACCATAAGGGAAAAAGTA	CTTATCATATCTTGTTAAATCA
247	TCTGCTAAACAATTAATTGGT	TTGCCATGGGTTATAG
248	TCTTTGATGGTGTTGTTATTC	AGAATTAATAATTTTCATGC
248d	TCTAAAACCTATTTGTCAAATG	AGAATTAATAATTTTCATGC
249	TGGGCTTACCATACTG	TTTTTTAGATGTTTTATGTG
250	TCTGGCCTTAATCTTAAGC	CTCTTTTACTTTAGCTTCA
251	TCTCAATATTTTTTGAAACAAG	TTTCAAACCTCCAGCCA
252	TTTATTTTCAGGTTATATCAA	GGAGTGCCTTTCTACT
253	TCTGAAAATTGGAAGTTTGC	TTCATATCGTAAAGCATC
254	TCTATTGAAAAGGGAGTTG	ATCGTCAACCTTAACG
255	TCTATTGTTGGTAGAGAAATCA	TTTTACTTGACGTCTCAC
256	TATCATGTAAAAATTGATCA	GTCTTCCATTAATATTCCC
257	TCTGATTTTTTATACAAAGGAGG	CCAATTATTTTGAAGTTC
258	TCTGAACGTTATACAGATAAAATG	ATTTTTTTGAATAATATAATCC
259	TCTCTTTCTCGTAAAAAAGAG	TTTATTATCAGAAAAGGC
260	TCTACTCTTGCTTAGTTGTTTAT	ATTCAAAAAATTTTCAA
261	TCTATAAGAAAAGCTGAAAATC	CGAAACGTCAGGTAAG
262	TCTATAAAAAATGCTATAGCATA	ACTTATTTTTGATAATATTCTT
263	TCTCAGCCTTCTAAACTACTTC	ATCAGCATTTCTACGAA
264	TCTGATTTGTTTAGCATGTTG	ATGTAGACTCCTAATGATTT
265	TCTCTTGCTTCCCTGATTT	TTTACTGTTCCCTTCGC
266	TCTCATCAATCAAATCATTATC	GAGATTAATTTGATTATATTTT
267	TCTATCTTTATTATCGGACAA	AACATCATTTTCTCCC
268	TCTAAAGAATTTATTAAGAATGG	GTTGATAGTTCCAAAACG
269	TCTGCAGATGATGGTGGTT	TAAATGTGTTTCTACTAAATT
270	TTAAATGATGCAATAACAA	CATCAATAGCCGAGCTG
271	TTGCTGGATTATCCTC	TTTATTTTCCAAATGACA
272	TCTGTATTTATGGCAAATAAGA	TTCACCTCGGAGTTGGAG
272d	TCTATGAGTTCTCTGGAAGTT	TTCACCTCGGAGTTGGAG
273	TCTGGTGTCTCAACTCTG	AATGTAAATGACAAAGGTA
274	TCTGTTTCATGATTTTGGTGA	GTTTTTTAATGGTTTGC
275	TCTGGGGTTTGGTTTTATA	TTTATCATAAGCATCTAGAC
276	TCTCAATCAGACATTAAAGCA	CTGATCTCTTGTTGATGC
277	TCTATTTGGAGGGGGGAAA	AAGCAGGGGAGCAATA
277d	TCTACCAAATTTGACTGGG	AAGCAGGGGAGCAATA
278	TCTGTTACGTTTTTCTTAT	CTGAGCAACACCTGTC
279	TCTAAAAAGAAAAGTTTAATTAGC	GGCAATTTTGTGGCAA
280	TTTGATTTTTTAAAGAAA	TTGCTTAGTTAATGGCT
281	TCTAAGAAATTAATTATAGGTATTT	AGGCGTTGAATATAATTC
281d	TCTGGTTTTTCGTTTTGA	AGGCGTTGAATATAATTC
282	TCTCTATTCTCAGATGAAACAA	CTTTTCAACTCCAAACA
283	TCTGTAAATTAATCGTTACTG	GAGTTGTCTTTTTTGTGTC
284	TCTATGCAACGATTAGGAC	GCAATCACAATTGACAT

285	TTAGGTGAAAGCAAATC	CTTTGTCTGCTTCACTT
286	TCTGGAGGATTTTATATGAAAG	TTGTATCTTCTCCTGACC
287	TCTGCACACACACCTACTAGT	TTGGTTAATCGTCTTG
287d	TCTAACAATCGTTCAAAGC	TTGGTTAATCGTCTTG
288	TCTAAAAAGTTTTTAAAAGTTTT	TTTAGTTACTTTCATAAATGG
288d	TGGAATAATCATCAGTCA	TTTAGTTACTTTCATAAATGG
289	TCTCAATCTAAAGGGCAAA	ATATAATTCTCTAAAAGTAGC
289L	TCTCAATCTAAAGGGCAAA	CCACTTCAAATTAACCTAAC
290	TATTACTTATCAAAAGAAAAGG	ATTCCTTGAACACGAA
291	TCTCAAGTATTAATGACAATGG	GTGCCATTCACTTCTCT
292	TTGAATCGTAAAAAAGG	TTGTCCTGTGAACCTGTG
293	TCTATGGGTCTAGCAACAA	AGGGTTTATTTGTTGAAG
293d N-term	TCTATGGGTCTAGCAACAA	TCCTGATTTATCCACTG
293d C-term	TCTGTTACAGCTAAACACGG	AGGGTTTATTTGTTGAAG
294	TCTGGTCATTTTAGTGAAAAA	CAAAATACCTAAGCTAGC
295	TCTAGCGACATAAAAATCAT	ACGAACCTCCATAACC
296	TCTAAAGGTATTATTTTAGCG	GGCTTCTCCAATCAAA
297	TCTATTCAGATTGGCAAATT	TTGAGTTAATGGATTGTT
298	TCTACTAAATTTATTGTTGATTCA	TAGCGTTATTTCACTGTG
299	TTTGAAATACTTAAACCTG	TTTCTCCGCCAGTCA
300	TCTGCTTCTACAAATAATGTTTC	CCGTTTATTTCTTCTACTG
301	TCTGTAATTAATATTGAGCAAGC	CATATCTGTTGCATCAAT
302	TCTGAAATCAACACTGAAATAG	AACTGGCTTTTAGTCAG
303	TCTACAAGGCATATAAAAATTTT	TTTATTATTTAATTCTTCAATA
304	TCTAACGAAATCAAATGCC	GTCTTTTAGAGCATCGA
305	TCTGGACGAGTAATGAAAACA	CTCTCCTCTAAGACTTTCCG
306	TCTGGGAAAAAATTGTTTT	TCCTTTTGTTACTTTTGC
307	TCTAAATTTACAGAACTTAACCTAT	TTTATCGCCTTTGTTG
308	ATGACACAGATGAATTTTA	ATGTTCAAGTTCTCCG
309	TTGCAACTTGGAATTG	TTCCATTATCTTCAAGTTA
310	TCTGCTAAAGAGAGGGTAGAT	CTCTTCTTCATTTTCTTA
311	TCAATTATTACTGATGTTTAC	TTTTTTTAAGTTGTAGAATG
312	TCTACTGCAACTAAACAACAT	GTTTTTTGATGCTTCTTG
313	TCTAAACGTATTGCTGTTTTA	TTTACTACTTTGGTTGGC
314	TCTAAATTTTATCTTGTTAGACAC	GTGTGTCATTTTGACCT
315	TCTATAGGGGATTATTCAGTAA	TCCTTCAAGATCATTTAA
316	TCTACTGAACGAACATTCTGA	ACCTCCTTTCTTTTCATT
317	TCTAATAAGCCATATTCAATAG	ATCTTCTCCTAACCTACCC
317d N-term	TCTAATAAGCCATATTCAATAG	ACTAGCTAGATTCTTAACGC
317d C-term	TCTGACTTGAATGGCAATAT	ATCTTCTCCTAACCTACCC
318	TCTATTGATTTTATTATTTCTATTG	GCCTCTTTCTCCAAAT
319	TTAAACATTTTGGTAGTAA	ATGTCCTGTTATATCTTCTT
320	TCTACTATTTATGACCAAATTG	GCGTTGAATAATGGTT
321	TCTAAAAATAAAAAAGATCAGTT	TATTTCTTTAGTTTCTTCAA
322	TCTCAAGAAACAGATACGACG	TAATAAAAAATTATATAAGAACCT
323	TCTGGTAATGAGTCAAAGAAC	TTCTGTCTTATAAGCATAAG
324	TCTGGAAGTAAATCAGCTTC	TTTTTTATAAGCATGTGTA
325	TCTGCTTGGCAACTTGTTTC	ATGAGACATAAGGTCTTG
326	TCTGGCATCTCAGACTTACC	GTTGGAGCTCCTACTG
326L	TCTAAATTCAAATCTGGGG	GTTGGAGCTCCTACTG
326L N-term	TCTAAATTCAAATCTGGGG	CATTTCTTTGGTTAAAGC
327	TCTGGAGGGAAAAATGAATC	TATCTCGAGTGCTATTTG
327d N-term	TCTGGAGGGAAAAATGAATC	CTCTTCATCGACATAGTAA
327d C-term	TCTGGCAACTTCAAAGCAT	TATCTCGAGTGCTATTTG
328	TCTGACCAAGTCGGTGTCC	ATTTTACAGTAGTGGAGTTT
329	TCTAAATCAAAGACCTCTTCTA	TGTCCTCATTTTTC
330	TCTAATAAACGCGTAAAAATC	TTTAACAGTACGAACACG

331	TCTACCAGAACAGTAGCAAT	CCCCCTGTTTTTAAAAAT
332	TCTACAAAAAACCTGTTATTAA	ACCCTCATATGATTCC
333	TCTATTGATATACAAAAATAAAA	TTTAAAAATAATGATACATCTC
333d	TCTGGATCATTGAGGGCAA	TTTAAAAATAATGATACATCTC
334	TCTAATTTAGTAAAAGTGAATAGTG	TAACCCCGTCTCAACA
335	TCTGAAGAAGAAAAATATTTTGA	TATTTTCGTTTTCTCAAA
336	TCTCAGGTTGAAGTTGACTTA	TTTCTCCAAATAATCTCTC
337	TCTGAAACAGATTCGTTTGTA	CCTACTTTTAGTTTTAGAAGA
338	TCTGCTATAATAGACAAAAAG	GAAATCATAGCTTCCC
339	TCGAAACCGATTAAGAT	ACCTTTTACTTTTGGTAGT
339d	TCTCAAGTCATGCGCTATG	ACCTTTTACTTTTGGTAGT
340	TCTGGATTTCTCTATAATTACTTC	TTGTTTGTGAAGTAAACG
341	TCTGGAAAACCATTTGTTAAC	TAATTTAAAAATTGCATAAA
342	TCTCAGAAAATTGAAGGTATT	TTTCGTTACCATATCTAGA
343	TCTGAAATGCAAGTTCAAA	TAAATCATGGAACTAGC
344	TCTGCACAACGCAGAATGT	AAAGCCCAACCTTCCG
345	TCTAAAAACCTGAATTGGG	GTTTCCACGTCCTTTC
346	TCTAATAAAATAGCTAATACAGAAG	AAGTTTATTCAAATCTGG
347	TCTATTGATATTCATTCTCATATC	AATGTAATGGTTTTTAATA
348	TCTACTGGATCTAAAAATTAGC	AGCTAAAATACCTAACCAG
349	TCTAAAGATCGCTTATATAATAAA	ATTTTTTAAACGACTCAT
350	TCTGCAAAAAGATATAATTAAGGTT	AGCGGAACGGTGAATA
351	TCAGAAGATCAAAAACA	ATAATCTAACTATCAGCTCT
352	TCTACTTTTTTTAAAAAGCTAAA	ATCTCCTATTGTAATTTTGA
352d	TCTGGTACAGATAGTAAATTTGG	ATCTCCTATTGTAATTTTGA
353	TCTACAATGTTAAAAATTGAAA	CACCTCTTTTGTGAGA
354	TCTATTAAAGAACTAAAAAGATTT	TTTGTTAGCGAGTAAGTC
355	TCTCGCTCACTACCTT	TTTATCATCCTCCTTAATAA
356	TCTAAATTCTATATTATTGATGATG	AAACGTTTTACTCTGTAAAA
357	TTGGAACATTTTTATATTAT	AAATAAGAATGTTAAAAAGAGC
358	TTTTATACAATTGAAGAGC	TTCCCCAAAAATTTCT
359	TCAAGAAATAATTACGGT	ACGCAGTCCCATTTC
360	TCTATAATGAAGGCGGTCT	CTGGCATGAGGTCTCA
361	TCTAGCGTATATGTTAGTGGA	CCTTTTTTCAATAATAGC
362	TCTACTAAACCACAGGGGG	ATCTTTAATCTTACCATCC
362d N-term	TCTACTAAACCACAGGGGG	TGCTGCTACTGCAATG
362 C-term	TCTGGTAATGAAGGAAATATCAC	ATCTTTAATCTTACCATCC
363	TCTCTCGAATTAAAAAATATTG	TAAATTCCTTTGTTGTAATA
364	TCTAACTATATGGGTATGGGC	ACCATCAGTTGTCACC
365	TCTGGAACTGCTACATATAGTAGG	TATTGACCAGTGCACG
366	TGGCTTGACATTATTTT	TTTTTTTGAATTTGTAAAAG
367	TCTAAGAAATTAAAAATATTCCT	AGAGATTATTTTATTTTAAAT
368	TCTAAAATCATTATTCACCGT	TTTATTTTTAGTATCTAAAACG
369	TCTAGTAGAATGATTCCAGG	TTTAGAACTCCAAGTATCTC
370	TCTACCGAATTTAATGACG	GTTAATTTGACTATTGATATATT
371	TCTAAAGATAGATATATTTTAGCAG	TAAACTCTCAAAAGCTAAAC
372	TCAGAAAAATATTCCACT	ACGTTCTTCTCTGGCT
373	TCTGAAATTGGTCAGCAAA	ACTTAAATGGAACAACC
374	TCTAAGTTGAAAAATATAATATATG	TTTGCTTAAAAAATTAGG
375	TCTGAAAAAGAACTATTTTAAGT	GGCTTTCCTCCCTTCA
376	TCTAAAGAAAAAGAAAAATTTGG	TTCATCTTTTTCAATATCA
377	TCTGGTAATAAACTGATGTATCA	GTGAGAGTGTCTTTGTTT
378	TCTGAAGATCAACTCACTATATTT	CAGATTTTTAGCTACTTGTC
379	TCTCAAATTACCCGAGAAG	TCTAGAGCGCTTTATAAG
380	TCTCTTAAAAGATTACTTACTGAAG	TTTTCTAATAGTTAGAAGCC
381	TCTCTTGGGATAGCTCACA	TTTTAAATGTGCAGAGA
382	TCTATAAAGTTTAAATTATTTTTTAA	ATTATAATTTCTTGGG

383	TCTATTTTACAGACGAATATACTAT	TCTATAATATCTCTCTAAAGTGA
384	TCTAGAATAATTGTTGTCGG	CCTCGCTAACATATCAC
385	TCTAATGTAAAAAACGC	AGCTCTTACAGTCTTGC
386	TCTCTAGTATCAAAGGAGAAAGC	TTGTCTGAGTGACCAA
387	TCTGGTATGTTGTTAGCA	ATAATATGAAATATGTTGTTCA
388	TCTCTTATGATAATAAATTCATTG	TCCGCAGAGTAAAAAA
389	TCTATGAATAGTGAACATAAAATT	TTCATAAATGTGCCAA
390	TCTAGGGAACTTACTGGA	TTCATCTCTGCTCACC
391	TCTAAAAAAGTCATCGATTTAA	TTCTCCTTCAGCTTTTA
392	TCTATTACATATGATTTACACAAG	GTCATTTTTTCTAAAGTTTG
393	TCTAATAAATCTTGTTGAGAA	TTTTGTAGTTGTTTCAAT
394	TCTCCTATGTTGTCTGTTGG	TTTCATTAGATAACTATTCCAGC
395	TCTACTTATCAAAAAACAGTTG	TATAGACTGAAGATAATTAATTA
396	TTTGTCAAAGGGATTT	AAATCGATTAAATCAAGTC
397	TCTAAATTATTTGATAAGTTTATAGA	TCTAAAGTAGTCCTTTAGACTA
397d	TCTAAAACTGCTACAGTTAG	TCTAAAGTAGTCCTTTAGACTA
398	TATTTAGAACAAATTAAGAGAGG	TTTGTCCATAATCATTTT
399	TCTAAAGTTTTAGTAGTTGATGAT	GGTAGATATGCCTAACATT
400	TCTAAAATAGTTGAAGGCG	GTTTCCTTCCAAAAAA
401	TCTGGAATTGAATTTAAAAATG	TCCATGCTTAATAGCC
402	TCTGGAAAATATTTTGGTACAG	ATCTAAACCAATTTCTGTAC
403	TCTGAGGTTAGAATGGTAAGTC	GTCCACAAAAACGTCT
404	TCTAAAATAGATGACCTAAGAAA	TAGATGTTCTACGGAGAA
405	TTGAAAATTCAGTATTATCA	AAAGATGGCAAGCCAT
406	TCTGATAAAAAATAATTTAGAAGACT	TCTCTCTCCACACCATA
407	TCTAAAATTGACATGAGGAA	CTTACCTCCTGTGGCT
407d	TCTAAAATTGACATGAGGAA	CTTTTGTGGTTACCTC
408	TCTAACCCTTACTTAACCTCA	TATTGTTAAATATGATGAAATG
409	TCTAAGGTAGTAGTAGCTATTGAT	ATGATTATACAAATTGATTAAT
409d	TCTACTGAAGAGAGAAATCCT	ATGATTATACAAATTGATTAAT
410	TCTGCTTTATTATCAGTTATTGTC	TCCCTCTTCCTTGACA
411	TCTAAAGACTATATTAACAGAATATT	AACGTTTTTGAGCTTT
412	TCTGGATTTTTTGACACAGC	TTTTGTCTTAAACGTTCT
413	TCTATTGTTGGTGAACAAGA	TTTAGATAGTCTAGCCATTT
414	TTAAATCAATATTTTCTGC	ACGGCTTGGGGCAGAG
415	TCTGAGCGAATTCCTGTTC	TACCATTATCCGTGCT
416	TCTGAAGTCATTTCGTGAACA	ACTATTAAACTCCAATGTTA
417	TCAAAACAATATGATTATATC	GCGCATTGTAACAAAT
418	TCTAGCAAGCCTAATGTTG	TTTTGGTAAAGGTCTG
419	TCTGATTTAAATAATTACATCGC	TCCTGGAAAGTTTCATC
420	TCTAAACGTGAATTACTACTCG	TAGTTTATCTAAAGCGTTC
421	TCTATACGCCAGTTTTTAAG	TTTATGTATAGAAACAGCAG
422	TTTTCGAGCGATTTTG	AATGTACATAACAATAGAGAGC
423	TCTGTAACCAAAGTTGAAGAG	CAACGATCCCAAGAAC
424	TCTATGAAAGATTTTATTGAATG	GCCATTCTTACCTCCT
424d	TCTATGAAAGATTTTATTGAATG	ACGTTTTTTCTGACCG
425	TCTATAGCCTTTAATAGTTTATTT	TATAAAATAAATTTGAAGATCT
426	TCTD440ACAGTTTATAATATAAACCATG	ATCATCTTGTACCAACTC
427	TATTCTTTTGAAGAACTTTT	GCCAATAAATTCACGG
428	TCTATAAAAAATTTTGATCCC	AGTCTGTTTTTAAACAAAAG
429	TCTAATCATTCCATTGAATC	TGGTTTTAGAACAACCTTTA
430	TTACAAAAAAAATATCGG	AATTAAGCTGAAAATGAC
431	TCTGCGGCTCAATTAGCTG	ATTATATTCTTTAATTTGTCA
432	TCTCGTACCTTCAAACCAG	CTTACGACGTCCTGGA
433	TCTATTAAAGCAACTTTTACTC	GTGTGTCATGACTACTGTAC
434	TCAATTTTTTCAGACAACA	TGAGTAGAGCACAAGC
642	TCTAGAAAACGTAATGATACATT	GAAACGAATACGTTCTT

-2990-

643	TCTGATTGTCAAATTACACCA	ACTACCTACCGTTTTTCAC
644	TCTATTTTTTCGTGGTGATAA	TTTGATGGTAACAGTCG
645	TTTTTTAATATTGAATATCAC	AGAAAGCGCTCTTCT
646	TCTAAGGGAGTCCAATATATG	TATCTTTAATAAAGCCCTA
647	TCTCGTCGCATGAATACCA	CATCCCATAAATTTGTT
648	TCTATAGAATTTTCAGGGC	CAAGACATTTCTTAAAGC
649	TCTGCTACTCACTCTAACTCAG	TTTTGTTTTAGCGATG
650	TGCTCTTCTTCAAATACT	TTTTAAACCATGCTGT
651	TCTCTAACACCATTACAAAAG	TTTGTAAGACCTTCTTT
652	TCTCAACAAGGTATTATGGATA	TTCTCGTTTATTAATTT
653	TCTAAAATTTTAGGTACACCA	AAAGAAAAGATGTGCC
654	TCTGGAAAAATGGTTAAGAA	CTGTGCAGGCTCAAAT
655	TCTAAATTCGTCCGAACCGT	AATTGTCCAGTCTAAGTTA
656	TCTGGTCTTCCAACGCAGC	ATTTAGTGTTATTTCTCCTG
657	TGCTCAGGTAAAACAT	TTTTTAAAGTGATGATGAA
658	TCTGAAAGCAAATCTTTGC	CTTTGTCTGCTTCACTT
659	TGTGCTAATTGGATTG	TTTTGGGGTTACTTTAC
660	TGTGGAAATGTCGGAG	TTTTGCTGAAATAATGTT
661	TGTCAGTCAAACCACA	ATCATACGAATGCAAC
662	TCTGCTAGTTTTATTTTTCC	TTTTTCATATTTTTTCAA
663	TGTGGAAGTAAATCAGC	ATTATTTTTATAAGCATGTG
664	TCTGTTAAATTAATAATCGTTACTG	GAGTTGTCTTTTTTGTG
665	TCTATTGCTGGTCTAGTG	GATAAGCACTTTCCTTAA
666	TTATTTTTTGAAATTGG	GCCTAAAAACCAATCA
667	TCTGCTGTATTTACACTCGTC	ATGTTTATGGCTTGCT
668	TTTTATATGAAAGACAACA	TTGTATCTTCTCCTGACC
669	TCAATTATTATTGGGTTAA	ATATACCCTAGACTTTTTGA
670	TCTCCTAAATTAACCCTAGTCT	GGCTTTAAAGTTCGATA
671	TCTAGTCTTGCGAAGGCAG	TTATCGTAAGCACTTAGG
672	TCTGTATTTACACTCGTCTTACA	ATGTTTATGGCTTGCTT
673	TCTGGAGGATTTTATATGAAAG	TTGTATCTTCTCCTGACC
674	TCTGTTAAATTAATAATCGTTACTG	GAGTTGTCTTTTTTGTCT
675	TCTGGTTCATCAGACAAACA	TTCAACTTGATTGCCA
676	TCTGTAGTTAAAGTTGGTATTAACG	TTTTGCAATTTTTGC
677	TCTGTATTAGAAGTACATGCTGA	TTTTAATGCTGTTTGAA
678	TCTGAGACACCAGTAATGGC	TTTTTTAGCTAAGGCTG
679	TCTGCTAACAAGCAGGATC	TTTTGCTAAACCTTCTG
680	TCTAATAAGTCCAGTAACTCTAAG	ATTCATATTAACACGATGC
681	TCTGCTTTTGATGTAATTATGC	TTTGCCTTTTGGAGGG
682	TCTATTAAGTATGAGGTTAAAGC	TGCACCTTGATGGCGA
683	TCTGTAATTGTTGAAGTTAGTTTG	CCATAATATTTGATGCTG
684	TCTCTTAGGAAGTATAAGCAA	TTCTAATCCTACAGCATG
685	TCTAAAATTTGTCTGGTTGG	AAAAATTCCTCCTAAATTA
686	TCTGACTTTTATGATATCAATCTT	AAAGTTTTGACTATTACTGATAG
687	TATGCTATTATGCAAAAAG	TGGGGGAGATAGTTATG
688	TCTGCAATCGTTTCAGCAG	TTGACAGAAAGCTAATTG

TABLE III – RESULTS FOR *in vivo* GBS CHALLENGE

GBS #	% survival	
	Pre-immune	Post-immune
1	18.7	22.2
4gst	19.4	37.2
4his	25.0	75.0
8	14.3	42.1
10	29.1	36.0
15	30.0	60.9

GBS #	% survival	
	Pre-immune	Post-immune
110	11.1	30.0
113	17.6	29.4
114	40.0	52.2
117	27.8	36.8
119	36.4	52.2
139	23.1	26.7

16	33.3	53.8	150	21.6	44.4
18	29.4	50.0	153	25.0	30.0
21	5.9	10.0	155	22.6	36.8
22	36.8	63.1	157	14.3	31.8
24	38.5	41.4	158	22.6	40.0
25	28.6	85.7	163	29.6	37.9
32	20.0	25.0	164	25.0	43.8
35	0.0	17.6	173	17.9	38.7
45	26.7	37.5	176	20.0	38.9
48	20.0	25.0	177	21.7	33.3
52	14.2	17.3	181	5.0	21.7
53	23.8	29.2	186	41.2	52.6
54	22.7	44.0	188	11.8	23.5
55	50.0	52.9	189	21.4	31.6
57	33.3	55.6	195	32.1	64.7
58	6.7	11.8	206	33.3	50.0
62	15.8	36.4	211	30.8	33.3
63	21.4	42.9	232	50.0	57.1
65	3.7	23.3	233	34.8	55.2
67	23.5	27.8	236	57.1	70.6
71	13.3	26.7	243	46.7	52.9
73	28.6	39.1	263	15.4	35.7
80	38.8	56.5	273	61.5	75.0
84	33.3	37.5	276	23.8	44.4
85	30.8	62.5	296	25.0	28.6
90	14.3	22.7	297	13.3	23.5
94	25.0	30.0	298	20.0	22.2
95	16.7	23.1	302	30.0	52.2
98	5.9	11.1	304	33.3	40.9
100	26.9	42.9	305	42.1	70.0
103	16.7	52.9	316	38.5	42.9
106	10.0	18.2	318	7.1	15.8

TABLE IV – COMPARISON OF GBS_{nnn} NUMBERING AND SEQ ID NUMBER

GBS numbering	Sequence listing	GBS numbering	Sequence listing
GBS1	SEQ ID 3532 & 8736	GBS345	SEQ ID 2442
GBS2	SEQ ID 4530 & 8818	GBS346	SEQ ID 2768
GBS3	SEQ ID 6266 & 8958	GBS347	SEQ ID 2766
GBS4	SEQ ID 2 & 8786	GBS348	SEQ ID 8658
GBS5	SEQ ID 2598 & 8674	GBS349	SEQ ID 2360
GBS6	SEQ ID 398 & 8496	GBS350	SEQ ID 8698
GBS7	SEQ ID 8790 & 9798	GBS351	SEQ ID 2970
GBS8	SEQ ID 8694	GBS352	SEQ ID 8692
GBS9	SEQ ID 4540 & 8822	GBS353	SEQ ID 3454
GBS10	SEQ ID 8718	GBS354	SEQ ID 8754
GBS11	SEQ ID 5884 & 8930	GBS355	SEQ ID 8752
GBS12	SEQ ID 8764 & 9692	GBS356	SEQ ID 8724
GBS13	SEQ ID 8484	GBS357	SEQ ID 8720
GBS14	SEQ ID 5406 & 8892	GBS358	SEQ ID 3184
GBS15	SEQ ID 4 & 8710	GBS359	SEQ ID 3948
GBS16	SEQ ID 944 & 8538	GBS360	SEQ ID 3926
GBS17	SEQ ID 1770 & 8602	GBS361	SEQ ID 8770
GBS18	SEQ ID 6860 & 9002	GBS362	SEQ ID 8768
GBS19	SEQ ID 4422 & 8812	GBS363	SEQ ID 3816
GBS20	SEQ ID 308 & 8488	GBS364	SEQ ID 1452
GBS21	SEQ ID 8762	GBS365	SEQ ID 1398

GBS22	SEQ ID 8584
GBS23	SEQ ID 8512
GBS24	SEQ ID 1694 & 8598
GBS25	SEQ ID 3180 & 8714
GBS26	SEQ ID 8820
GBS27	SEQ ID 8774
GBS28	SEQ ID 8738
GBS29	SEQ ID 8744
GBS30	SEQ ID 8860
GBS31	SEQ ID 8702
GBS32	SEQ ID 8910 & 10142
GBS33	SEQ ID 5734 & 8912
GBS34	SEQ ID 5750 & 8916
GBS35	SEQ ID 8908
GBS36	SEQ ID 8542
GBS37	SEQ ID 8564
GBS38	SEQ ID 2122 & 8642
GBS39	SEQ ID 8480
GBS40	SEQ ID 8654
GBS41	SEQ ID 1176 & 8562
GBS42	SEQ ID 4856 & 8850
GBS43	SEQ ID 672 & 8520
GBS44	SEQ ID 9000
GBS45	SEQ ID 9018
GBS46	SEQ ID 1834 & 8608
GBS47	SEQ ID 8588
GBS48	SEQ ID 8594 & 8596
GBS49	SEQ ID 8494 & 9490
GBS50	SEQ ID 1236 & 8566
GBS51	SEQ ID 5410
GBS52	SEQ ID 3920
GBS53	SEQ ID 8586
GBS54	SEQ ID 3442
GBS55	SEQ ID 9020 & 10338
GBS56	SEQ ID 2510 & 8668
GBS57	SEQ ID 8854
GBS58	SEQ ID 8664
GBS59	SEQ ID 3744
GBS60	SEQ ID 8760
GBS61	SEQ ID 8776
GBS62	SEQ ID 2244
GBS63	SEQ ID 390
GBS64	SEQ ID 374
GBS65	SEQ ID 8544
GBS66	SEQ ID 3028
GBS67	SEQ ID 3746
GBS68	SEQ ID 4012
GBS69	SEQ ID 4916
GBS70	SEQ ID 3718
GBS71	SEQ ID 8906
GBS72	SEQ ID 1348
GBS73	SEQ ID 220
GBS74	SEQ ID 5872
GBS75	SEQ ID 8926
GBS76	SEQ ID 5862
GBS77	SEQ ID 3256
GBS78	SEQ ID 3262
GBS79	SEQ ID 3264
GBS80	SEQ ID 8780

GBS366	SEQ ID 8574
GBS367	SEQ ID 1340
GBS368	SEQ ID 1598
GBS369	SEQ ID 4822
GBS370	SEQ ID 8844
GBS371	SEQ ID 4926
GBS372	SEQ ID 4956
GBS373	SEQ ID 5062
GBS374	SEQ ID 8878
GBS375	SEQ ID 326
GBS376	SEQ ID 5380
GBS377	SEQ ID 5468
GBS378	SEQ ID 5570
GBS379	SEQ ID 8918
GBS380	SEQ ID 156
GBS381	SEQ ID 8934
GBS382	SEQ ID 8610
GBS383	SEQ ID 4738
GBS384	SEQ ID 8836
GBS385	SEQ ID 1094
GBS386	SEQ ID 9038
GBS387	SEQ ID 8558
GBS388	SEQ ID 9040
GBS389	SEQ ID 8516
GBS390	SEQ ID 8952
GBS391	SEQ ID 8522
GBS392	SEQ ID 6220
GBS393	SEQ ID 8966
GBS394	SEQ ID 8960
GBS395	SEQ ID 6276
GBS396	SEQ ID 8468
GBS397	SEQ ID 6262
GBS398	SEQ ID 8806
GBS399	SEQ ID 1960
GBS400	SEQ ID 3154
GBS401	SEQ ID 3170
GBS402	SEQ ID 4236
GBS403	SEQ ID 8798
GBS404	SEQ ID 8800
GBS405	SEQ ID 8508
GBS406	SEQ ID 8506
GBS407	SEQ ID 6484
GBS408	SEQ ID 9042
GBS409	SEQ ID 6678
GBS410	SEQ ID 4064
GBS411	SEQ ID 9044
GBS412	SEQ ID 9046
GBS413	SEQ ID 272
GBS414	SEQ ID 8946
GBS415	SEQ ID 8944
GBS416	SEQ ID 6044
GBS417	SEQ ID 1874
GBS418	SEQ ID 5146
GBS419	SEQ ID 2638
GBS420	SEQ ID 2104
GBS421	SEQ ID 2108
GBS422	SEQ ID 714
GBS423	SEQ ID 6884
GBS424	SEQ ID 4874

GBS81	SEQ ID 2706
GBS82	SEQ ID 2898
GBS83	SEQ ID 8772
GBS84	SEQ ID 4182
GBS85	SEQ ID 216
GBS86	SEQ ID 2978
GBS87	SEQ ID 3452
GBS88	SEQ ID 5694
GBS89	SEQ ID 2682
GBS90	SEQ ID 8476
GBS91	SEQ ID 8938
GBS92	SEQ ID 8964 & 10238
GBS93	SEQ ID 2848
GBS94	SEQ ID 1592
GBS95	SEQ ID 2224
GBS96	SEQ ID 2130
GBS97	SEQ ID 800
GBS98	SEQ ID 8746
GBS99	SEQ ID 4240
GBS100	SEQ ID 8782
GBS101	SEQ ID 6902
GBS102	SEQ ID 6894
GBS103	SEQ ID 6
GBS104	SEQ ID 8778
GBS105	SEQ ID 1400
GBS106	SEQ ID 8502
GBS107	SEQ ID 6026
GBS108	SEQ ID 8532
GBS109	SEQ ID 4116
GBS110	SEQ ID 6832
GBS111	SEQ ID 8842
GBS112	SEQ ID 8904
GBS113	SEQ ID 300
GBS114	SEQ ID 8968
GBS115	SEQ ID 5164
GBS116	SEQ ID 5152
GBS117	SEQ ID 8962
GBS118	SEQ ID 2508
GBS119	SEQ ID 8814
GBS120	SEQ ID 8874
GBS121	SEQ ID 3826
GBS122	SEQ ID 9006
GBS123	SEQ ID 6310
GBS124	SEQ ID 260
GBS125	SEQ ID 3872
GBS126	SEQ ID 6736
GBS127	SEQ ID 8816
GBS128	SEQ ID 752
GBS129	SEQ ID 8990
GBS130	SEQ ID 9004
GBS131	SEQ ID 6198
GBS132	SEQ ID 8730
GBS133	SEQ ID 474
GBS134	SEQ ID 9008
GBS135	SEQ ID 8882
GBS136	SEQ ID 1188
GBS137	SEQ ID 3960
GBS138	SEQ ID 9052
GBS139	SEQ ID 884

GBS425	SEQ ID 3978
GBS426	SEQ ID 3976
GBS427	SEQ ID 6958
GBS428	SEQ ID 3398
GBS429	SEQ ID 3402
GBS430	SEQ ID 8840
GBS431	SEQ ID 8902
GBS432	SEQ ID 8534
GBS433	SEQ ID 2558
GBS434	SEQ ID 8590
GBS435	SEQ ID 484
GBS436	SEQ ID 8472
GBS437	SEQ ID 466
GBS438	SEQ ID 362
GBS439	SEQ ID 900
GBS440	SEQ ID 8536
GBS441	SEQ ID 936
GBS442	SEQ ID 940
GBS443	SEQ ID 998
GBS444	SEQ ID 1776
GBS445	SEQ ID 8634
GBS446	SEQ ID 2048
GBS447	SEQ ID 1654
GBS448	SEQ ID 8592
GBS449	SEQ ID 1634
GBS450	SEQ ID 1630
GBS451	SEQ ID 2098
GBS452	SEQ ID 2062
GBS453	SEQ ID 8636
GBS454	SEQ ID 1734
GBS455	SEQ ID 1690
GBS456	SEQ ID 1684
GBS457	SEQ ID 8656
GBS458	SEQ ID 8650
GBS459	SEQ ID 2152
GBS460	SEQ ID 2148
GBS461	SEQ ID 2394
GBS462	SEQ ID 2778
GBS463	SEQ ID 8688
GBS464	SEQ ID 8684
GBS465	SEQ ID 8682
GBS466	SEQ ID 2694
GBS467	SEQ ID 2350
GBS468	SEQ ID 8660
GBS469	SEQ ID 2998
GBS470	SEQ ID 2988
GBS471	SEQ ID 2924
GBS472	SEQ ID 2910
GBS473	SEQ ID 2882
GBS474	SEQ ID 2878
GBS475	SEQ ID 2856
GBS476	SEQ ID 8690
GBS477	SEQ ID 3112
GBS478	SEQ ID 3432
GBS479	SEQ ID 3460
GBS480	SEQ ID 3504
GBS481	SEQ ID 8734
GBS482	SEQ ID 8740
GBS483	SEQ ID 3606

GBS140	SEQ ID 8632
GBS141	SEQ ID 1768
GBS142	SEQ ID 8600
GBS143	SEQ ID 9054
GBS144	SEQ ID 2238
GBS145	SEQ ID 8700
GBS146	SEQ ID 8696
GBS147	SEQ ID 8526
GBS148	SEQ ID 9010
GBS149	SEQ ID 8732
GBS150	SEQ ID 3736
GBS151	SEQ ID 3188
GBS152	SEQ ID 3952
GBS153	SEQ ID 3904
GBS154	SEQ ID 4024
GBS155	SEQ ID 8796
GBS156	SEQ ID 4646
GBS157	SEQ ID 4812
GBS158	SEQ ID 5504
GBS159	SEQ ID 8628
GBS160	SEQ ID 8924
GBS161	SEQ ID 8922
GBS162	SEQ ID 168
GBS163	SEQ ID 224
GBS164	SEQ ID 1102
GBS165	SEQ ID 3672
GBS166	SEQ ID 8712
GBS167	SEQ ID 4214
GBS168	SEQ ID 9016
GBS169	SEQ ID 4346
GBS170	SEQ ID 8982
GBS171	SEQ ID 6720
GBS172	SEQ ID 6704
GBS173	SEQ ID 8788
GBS174	SEQ ID 6150
GBS175	SEQ ID 62
GBS176	SEQ ID 8478
GBS177	SEQ ID 8876
GBS178	SEQ ID 6078
GBS179	SEQ ID 8848
GBS180	SEQ ID 3062
GBS181	SEQ ID 1924
GBS182	SEQ ID 3774
GBS183	SEQ ID 4796
GBS184	SEQ ID 1978
GBS185	SEQ ID 1046
GBS186	SEQ ID 8470
GBS187	SEQ ID 844
GBS188	SEQ ID 3410
GBS189	SEQ ID 6986
GBS190	SEQ ID 8842
GBS191	SEQ ID 1814
GBS192	SEQ ID 8618
GBS193	SEQ ID 2382
GBS194	SEQ ID 3912
GBS195	SEQ ID 8
GBS196	SEQ ID 4944
GBS197	SEQ ID 5486
GBS198	SEQ ID 8896

GBS484	SEQ ID 3562
GBS485	SEQ ID 3552
GBS486	SEQ ID 3762
GBS487	SEQ ID 3756
GBS488	SEQ ID 3732
GBS489	SEQ ID 3730
GBS490	SEQ ID 3704
GBS491	SEQ ID 3698
GBS492	SEQ ID 3252
GBS493	SEQ ID 3244
GBS494	SEQ ID 3238
GBS495	SEQ ID 8722
GBS496	SEQ ID 8716
GBS497	SEQ ID 3876
GBS498	SEQ ID 3858
GBS499	SEQ ID 8758
GBS500	SEQ ID 4022
GBS501	SEQ ID 4106
GBS502	SEQ ID 1406
GBS503	SEQ ID 8580
GBS504	SEQ ID 4578
GBS505	SEQ ID 4566
GBS506	SEQ ID 8832
GBS507	SEQ ID 8830
GBS508	SEQ ID 4644
GBS509	SEQ ID 8828
GBS510	SEQ ID 8826
GBS511	SEQ ID 4892
GBS512	SEQ ID 4970
GBS513	SEQ ID 4974
GBS514	SEQ ID 8862
GBS515	SEQ ID 8864
GBS516	SEQ ID 8866
GBS517	SEQ ID 8868
GBS518	SEQ ID 9012
GBS519	SEQ ID 5068
GBS520	SEQ ID 8870
GBS521	SEQ ID 5228
GBS522	SEQ ID 322
GBS523	SEQ ID 8492
GBS524	SEQ ID 8894
GBS525	SEQ ID 5430
GBS526	SEQ ID 5414
GBS527	SEQ ID 5524
GBS528	SEQ ID 8898
GBS529	SEQ ID 5670
GBS530	SEQ ID 5630
GBS531	SEQ ID 5588
GBS532	SEQ ID 1324
GBS533	SEQ ID 8914
GBS534	SEQ ID 8550
GBS535	SEQ ID 8568
GBS536	SEQ ID 1288
GBS537	SEQ ID 5798
GBS538	SEQ ID 8920
GBS539	SEQ ID 158
GBS540	SEQ ID 8482
GBS541	SEQ ID 184
GBS542	SEQ ID 9048

GBS199	SEQ ID 1162
GBS200	SEQ ID 8936
GBS201	SEQ ID 4550
GBS202	SEQ ID 8666
GBS203	SEQ ID 6478
GBS204	SEQ ID 1996
GBS205	SEQ ID 18
GBS206	SEQ ID 8552
GBS207	SEQ ID 3822
GBS208	SEQ ID 3916
GBS209	SEQ ID 3918
GBS210	SEQ ID 3738
GBS211	SEQ ID 4680
GBS212	SEQ ID 8750
GBS213	SEQ ID 8500
GBS214	SEQ ID 8498
GBS215	SEQ ID 9022
GBS216	SEQ ID 8606
GBS217	SEQ ID 9024
GBS218	SEQ ID 8652
GBS219	SEQ ID 8646
GBS220	SEQ ID 2730
GBS221	SEQ ID 9028
GBS222	SEQ ID 3842
GBS223	SEQ ID 8794
GBS224	SEQ ID 9026
GBS225	SEQ ID 8834
GBS226	SEQ ID 4966
GBS227	SEQ ID 5030
GBS228	SEQ ID 5050
GBS229	SEQ ID 9056
GBS230	SEQ ID 1296
GBS231	SEQ ID 5810
GBS232	SEQ ID 5830
GBS233	SEQ ID 4722
GBS234	SEQ ID 1106
GBS235	SEQ ID 8560
GBS236	SEQ ID 6162
GBS237	SEQ ID 8706
GBS238	SEQ ID 4246
GBS239	SEQ ID 8980
GBS240	SEQ ID 8986
GBS241	SEQ ID 9030
GBS242	SEQ ID 9032
GBS243	SEQ ID 8678
GBS244	SEQ ID 6554
GBS245	SEQ ID 8994
GBS246	SEQ ID 6864
GBS247	SEQ ID 8856
GBS248	SEQ ID 454
GBS249	SEQ ID 8620
GBS250	SEQ ID 8634
GBS251	SEQ ID 2258
GBS252	SEQ ID 8648
GBS253	SEQ ID 2526
GBS254	SEQ ID 2710
GBS255	SEQ ID 2966
GBS256	SEQ ID 3424
GBS257	SEQ ID 3550

GBS543	SEQ ID 8932
GBS544	SEQ ID 5880
GBS545	SEQ ID 44
GBS546	SEQ ID 9014
GBS547	SEQ ID 12
GBS548	SEQ ID 8614
GBS549	SEQ ID 8612
GBS550	SEQ ID 4720
GBS551	SEQ ID 4710
GBS552	SEQ ID 1086
GBS553	SEQ ID 1088
GBS554	SEQ ID 1138
GBS555	SEQ ID 8748
GBS556	SEQ ID 5968
GBS557	SEQ ID 774
GBS558	SEQ ID 1192
GBS559	SEQ ID 1196
GBS560	SEQ ID 1268
GBS561	SEQ ID 8518
GBS562	SEQ ID 8676
GBS563	SEQ ID 2296
GBS564	SEQ ID 2300
GBS565	SEQ ID 8950
GBS566	SEQ ID 694
GBS567	SEQ ID 680
GBS568	SEQ ID 6300
GBS569	SEQ ID 8956
GBS570	SEQ ID 8972
GBS571	SEQ ID 8970
GBS572	SEQ ID 3300
GBS573	SEQ ID 3304
GBS574	SEQ ID 8726
GBS575	SEQ ID 8810
GBS576	SEQ ID 4418
GBS577	SEQ ID 8808
GBS578	SEQ ID 4382
GBS579	SEQ ID 4378
GBS580	SEQ ID 1932
GBS581	SEQ ID 8622
GBS582	SEQ ID 8624
GBS583	SEQ ID 1962
GBS584	SEQ ID 8708
GBS585	SEQ ID 8672
GBS586	SEQ ID 6444
GBS587	SEQ ID 8976
GBS588	SEQ ID 8804
GBS589	SEQ ID 8514
GBS590	SEQ ID 8510
GBS591	SEQ ID 630
GBS592	SEQ ID 8504
GBS593	SEQ ID 514
GBS594	SEQ ID 8978
GBS595	SEQ ID 6738
GBS596	SEQ ID 6712
GBS597	SEQ ID 6686
GBS598	SEQ ID 6674
GBS599	SEQ ID 6662
GBS600	SEQ ID 8988
GBS601	SEQ ID 8578

GBS258	SEQ ID 3752
GBS259	SEQ ID 8756
GBS260	SEQ ID 4162
GBS261	SEQ ID 1530
GBS262	SEQ ID 8572
GBS263	SEQ ID 1616
GBS264	SEQ ID 8824
GBS265	SEQ ID 4554
GBS266	SEQ ID 4652
GBS267	SEQ ID 4980
GBS268	SEQ ID 5038
GBS269	SEQ ID 5534
GBS270	SEQ ID 1998
GBS271	SEQ ID 8570
GBS272	SEQ ID 22
GBS273	SEQ ID 5994
GBS274	SEQ ID 774
GBS275	SEQ ID 2308
GBS276	SEQ ID 8942
GBS277	SEQ ID 8954
GBS278	SEQ ID 8524
GBS279	SEQ ID 6292
GBS280	SEQ ID 6254
GBS281	SEQ ID 4458
GBS282	SEQ ID 4444
GBS283	SEQ ID 9034
GBS284	SEQ ID 6456 & 8974
GBS285	SEQ ID 8802
GBS286	SEQ ID 9036
GBS287	SEQ ID 5354
GBS288	SEQ ID 5374
GBS289	SEQ ID 8616
GBS290	SEQ ID 8680
GBS291	SEQ ID 8530
GBS292	SEQ ID 8998
GBS293	SEQ ID 8582
GBS294	SEQ ID 8604
GBS295	SEQ ID 2722
GBS296	SEQ ID 2658
GBS297	SEQ ID 3024
GBS298	SEQ ID 8704
GBS299	SEQ ID 3268
GBS300	SEQ ID 4170
GBS301	SEQ ID 8576
GBS302	SEQ ID 8670
GBS303	SEQ ID 8554
GBS304	SEQ ID 5846
GBS305	SEQ ID 208
GBS306	SEQ ID 212
GBS307	SEQ ID 8992
GBS308	SEQ ID 8880
GBS309	SEQ ID 3386
GBS310	SEQ ID 286
GBS311	SEQ ID 3964
GBS312	SEQ ID 4660
GBS313	SEQ ID 4090
GBS314	SEQ ID 8556
GBS315	SEQ ID 1766
GBS316	SEQ ID 2000

GBS602	SEQ ID 8948
GBS603	SEQ ID 6132
GBS604	SEQ ID 5282
GBS605	SEQ ID 5302
GBS606	SEQ ID 8884
GBS607	SEQ ID 5314
GBS608	SEQ ID 8886
GBS609	SEQ ID 8888
GBS610	SEQ ID 8890
GBS611	SEQ ID 6028
GBS612	SEQ ID 8474
GBS613	SEQ ID 5092
GBS614	SEQ ID 8872
GBS615	SEQ ID 6052
GBS616	SEQ ID 8940
GBS617	SEQ ID 1824
GBS618	SEQ ID 6600
GBS619	SEQ ID 6608
GBS620	SEQ ID 6620
GBS621	SEQ ID 864
GBS622	SEQ ID 8640
GBS623	SEQ ID 8996
GBS624	SEQ ID 9050
GBS625	SEQ ID 2812
GBS626	SEQ ID 8858
GBS627	SEQ ID 8852
GBS628	SEQ ID 8784
GBS629	SEQ ID 6950
GBS630	SEQ ID 4502
GBS631	SEQ ID 4492
GBS632	SEQ ID 4488
GBS633	SEQ ID 8728
GBS634	SEQ ID 3066
GBS635	SEQ ID 8838
GBS636	SEQ ID 4772
GBS637	SEQ ID 8626
GBS638	SEQ ID 8984
GBS639	SEQ ID 8546
GBS640	SEQ ID 6780
GBS641	SEQ ID 900
GBS642	1312
GBS643	1772
GBS644	1956
GBS645	2726
GBS646	3348
GBS647	3770
GBS648	4934
GBS649	5076
GBS650	5446
GBS651	5602
GBS652	5610
GBS653	5760
GBS654	6096
GBS655	6656
GBS656	9324
GBS657	10782
GBS658	8802
GBS659	9344
GBS660	9410

GBS317	SEQ ID 4210	GBS661	9428
GBS318	SEQ ID 8548	GBS662	9286
GBS319	SEQ ID 892	GBS663	9294
GBS320	SEQ ID 916	GBS664	9034
GBS321	SEQ ID 8846	GBS665	10546
GBS322	SEQ ID 8540	GBS666	10610
GBS323	SEQ ID 2102	GBS667	9052
GBS324	SEQ ID 8490	GBS668	9036
GBS325	SEQ ID 8900	GBS669	9010
GBS326	SEQ ID 8630	GBS670	10730
GBS327	SEQ ID 5856	GBS671	9020
GBS328	SEQ ID 6016	GBS672	9052
GBS329	SEQ ID 8928	GBS673	9036
GBS330	SEQ ID 8792	GBS674	9034
GBS331	SEQ ID 922	GBS675	10634
GBS332	SEQ ID 1004	GBS676	10692
GBS333	SEQ ID 1786	GBS677	10746
GBS334	SEQ ID 1784	GBS678	9330
GBS335	SEQ ID 1782	GBS679	9404
GBS336	SEQ ID 1886	GBS680	6668
GBS337	SEQ ID 2010	GBS681	4264
GBS338	SEQ ID 8638	GBS682	6762
GBS339	SEQ ID 2080	GBS683	9290
GBS340	SEQ ID 8594 & 8596	GBS684	9614
GBS341	SEQ ID 2280	GBS685	10454
GBS342	SEQ ID 2266	GBS686	2774
GBS343	SEQ ID 8644	GBS687	4620
GBS344	SEQ ID 8662	GBS688	10224

TABLE V – NUCLEOTIDES DELETED IN EXPRESSION OF GBS_{nnn} PROTEINS

GBS	Deleted nucleotides	GBS	Deleted nucleotides
11d	1-153	272d	1-531
31d	1-129	277d	1-318
64d	1-165	281d	1-54
68d	2029-2796	287d	1-108
70d	1-402	288d	1-72
74d	1-975	293C	1-1229
79d	1-201	293N	1230-2379
105dN	2689-4119	317N	1729-4107
105dC	1-2688	317C	1-2379
105d	1-2688	326N	1707-2652
109d	1-120	326dN	2326-3927
130d	1-518	327N	3034-6831
170d	1-111	327C	1-3033
182d	1596-1674	333d	1-150
195C	1-1710	339d	1-111
195N	1711-3243	352d	1-158
209d	757-912	362N	1707-2652
210d	1-99 & 777-879	362C	1-1706
220d	1-120	397d	1-348
231d	1-54	399d	1-111
235d	1-270	407d	1174-1473
246d	1-75	409d	1-297
248d	1-591	424d	1327-1671

TABLE VI – PREDICTED FUNCTIONS FOR CERTAIN SEQ IDs

SEQ ID	Function
6	manganese ABC transporter, ATP-binding protein (psaB)
12	iron (chelated) ABC transporter, permease protein (psaC)
18	peptidyl-prolyl cis-trans isomerase, cyclophilin-type
26	chorismate binding enzyme (pabB)
30	probable transposase (insertion sequence IS861)
42	peptidase, M20/M25/M40 family
44	drug transporter
50	ribosomal protein L11 (rplK)
54	ribosomal protein L1 (rplA)
62	peptide ABC transporter, permease protein
66	peptide ABC transporter, permease protein
78	uridylate kinase (pyrH)
84	ribosome recycling factor (frr)
104	PhoH family protein (phoH)
110	MutT/nudix family protein superfamily
116	tetracenomycin polyketide synthesis O-methyltransferase TcmP
134	phosphopantetheine adenylyltransferase (coaD)
140	PDZ domain protein
144	5-nucleotidase family protein
156	VanZF-related protein
158	ABC transporter, ATP-binding/permease protein
162	ABC transporter, ATP-binding/permease protein
168	BioY family protein
180	acetyl-CoA acetyltransferase
188	endonuclease III (nth)
196	glucokinase (gki)
200	rhodanese family protein
204	elongation factor Tu family protein (typA)
212	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-
216	cell division protein DivIB
220	cell division protein FtsA (ftsA)
224	cell division protein FtsZ (ftsZ)
236	ylmH protein (ymH)
240	cell division protein DivIVA (divIVA)
244	isoleucyl-tRNA synthetase (ileS)
252	MutT/nudix family protein
256	ATP-dependent Clp protease, ATP-binding subunit ClpE (clpE)
268	methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclo-
274	exodeoxyribonuclease VII, large subunit (xseA)
278	exodeoxyribonuclease VII, small subunit (xseB)
282	geranyltranstransferase (ispA)
286	hemolysin A
290	transcriptional repressor
296	DNA repair protein RecN (recN)
300	degV family protein (degV)
322	peptide ABC transporter, permease protein (oppC)
326	peptide ABC transporter, ATP-binding protein (oppD)
328	peptide ABC transporter, ATP-binding protein (oppF)
348	4-diphosphocytidyl-2C-methyl-D-erythritol kinase (ispE)
352	adc operon repressor AdcR (adcR)
356	zinc ABC transporter, ATP-binding protein (adcC)
370	tyrosyl-tRNA synthetase (tyrS)
374	penicillin-binding protein 1B (pbp1B)
378	DNA-directed RNA polymerase, beta subunit (rpoB)
382	dna-directed rna polymerase beta' chain

-2999-

390	competence protein CglA (cglA)
406	acetate kinase (ackA)
410	transcriptional regulator
418	pyrroline-5-carboxylate reductase (proC)
422	glutamyl-aminopeptidase (pepA)
432	thioredoxin family protein
436	tRNA binding domain protein (pheT)
440	methyltransferase
442	single-strand DNA-binding protein, authentic point mutation (ssbB)
454	GAF domain protein (lytS)
466	lrgB protein (lrgB)
474	oligopeptide ABC transporter, permease protein
476	peptide ABC transporter, ATP-binding protein
480	peptide ABC transporter, ATP-binding protein (oppF)
484	PTS system, IIA _{BC} components (treB)
488	alpha amylase family protein (treC)
494	transcriptional regulator, BglG family
506	transcriptional regulator, BglG family
508	PTS system, IIB component
514	PTS system, IIC component
518	transketolase, N-terminal subunit (tktA)
528	ribosomal protein S15 (rpsO)
546	cysteinyI-tRNA synthetase (cysS)
554	RNA methyltransferase, TrmH family, group 3
562	DegV family protein (degV)
572	ribosomal protein S9 (rpsI)
576	integrase, phage family
580	transcriptional regulator
596	recombination protein
626	transcriptional regulator MutR
630	transporter
640	amino acid ABC transporter, permease protein (opuBB)
642	glycine betaine/L-proline transport ATP binding subunit (proV)
654	lectin, alpha subunit precursor
662	transcriptional regulator
664	acetyltransferase, GNAT family
666	acetyltransferase, GNAT family (rimJ)
670	acetyltransferase, GNAT family
676	transcriptional regulator, tetR family domain protein
680	ABC transporter efflux protein, DrrB family
690	IS1381, transposase OrfA/OrfB, truncation
714	magnesium transporter, CorA family
718	oxidoreductase, Gfo/Idh/MocA family
722	valyl-tRNA synthetase (valS)
730	acetyltransferase, GNAT family
746	methyltransferase
750	bacteriophage L54a, integrase
754	DNA-damage-inducible protein J
774	cation efflux system protein
778	oxidoreductase, aldo/keto reductase family
784	alcohol dehydrogenase, zinc-containing
790	3-oxoadipate enol-lactone hydrolase/4-carboxymuconolactone decarboxylase
804	ribonucleoside-diphosphate reductase, alpha subunit (nrdE)
808	nrdI protein (nrdI)
812	Ribonucleotide reductases
824	elaA protein (elaA)
828	RNA methyltransferase, TrmA family
832	RecX family protein
840	-identity (jag)

844	membrane protein, 60 kDa (yidC)
856	UTP-glucose-1-phosphate uridylyltransferase (galU)
864	rhomboid family protein
884	MORN motif family
892	transcriptional regulator
896	adenylosuccinate lyase (purB)
908	phosphoribosylaminoimidazole carboxylase, catalytic subunit (purE)
912	phosphoribosylamine-glycine ligase (purD)
916	phosphosugar-binding transcriptional regulator
920	acetyl xylan esterase
922	ROK family protein (gki)
926	N-acetylneuraminate lyase (nanA)
936	sugar ABC transporter, permease protein
940	sugar ABC transporter, permease protein (msmF)
952	LysM domain protein, authentic frameshift
956	zoocin A endopeptidase
958	phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydr
962	acetyltransferase, GNAT family family
964	phosphoribosylglycinamide formyltransferase (purN)
968	phosphoribosylformylglycinamide cyclo-ligase (purM)
972	amidophosphoribosyltransferase (purF)
980	phosphoribosylformylglycinamide synthase
984	phosphoribosylaminoimidazole-succinocarboxamide synthase (purC)
1042	oligoendopeptidase F (pepF)
1060	ebsC protein
1068	hydrolase, haloacid dehalogenase-like family
1076	riboflavin synthase, beta subunit (ribH)
1082	riboflavin biosynthesis protein RibD (ribD)
1086	Mn2+/Fe2+ transporter, NRAMP family
1094	peptidase, U32 family
1116	HPr(Ser) kinase/phosphatase (hprK)
1130	oxidoreductase
1148	signal recognition particle-docking protein FtsY (ftsY)
1152	Cof family protein
1156	Cof family protein
1172	vicX protein (vicX)
1176	sensory box sensor histidine kinase (vicK)
1180	DNA-binding response regulator (vicR)
1184	amino acid ABC transporter, ATP-binding protein
1188	amino acid ABC transporter, amino acid-binding protein (fliY)
1192	amino acid ABC transporter, permease protein
1196	amino acid ABC transporter, permease protein
1208	DNA-binding response regulator (vicR)
1210	threonyl-tRNA synthetase (thrS)
1214	glycosyl transferase, group 1
1218	glycosyl transferase, group 1 (cpoA)
1222	alpha-amylase (amy)
1230	proline dipeptidase (pepQ)
1238	haloacid dehalogenase-like hydrolase superfamily
1244	mannonate dehydratase (uxuA)
1248	glucuronate isomerase
1254	transcriptional regulator, GntR family
1268	sodiumgalactoside symporter family protein
1270	D-isomer specific 2-hydroxyacid dehydrogenase family protein
1282	transcriptional regulator, LysR family
1290	ABC transporter, ATP-binding protein (potA)
1296	DedA family protein
1308	MutT/nudix family protein family
1310	phosphoserine phosphatase SerB (serB)